# *PREVENT* AND *STOP* COMPLEMENTATION CLAUSES: A CORPUS-BASED INVESTIGATION OF 19TH, 20TH AND 21ST CENTURY AMERICAN ENGLISH

by

TERESA WAI SEE ONG

A thesis submitted to
The University of Birmingham
for the degree of
MASTER OF PHILOSOPHY

Department of English
School of English, Drama, and American and Canadian Studies
College of Arts and Law
University of Birmingham
September 2011

# ABSTRACT

This study is a corpus-based investigation into the diachronic development of non-finite complementation clauses. My aim is to find out how the complementation clauses of two semantically comparable verbs, *prevent* and *stop* have changed over the past 200 years in American English. Two specific variants were considered:

(i) noun phrase + *from* + *-ing* (e.g. She *prevented / stopped it from eating.*)

(ii) noun phrase + *-ing* (e.g. She *prevented / stopped it eating.*)

All the complementation clauses were extracted from the 400 million word Corpus of Historical American English. The *prevent* and *stop* complementation clauses were examined in respect of quantitative changes from 1810 to 2009, with focus on the overall frequencies and proportional values. The length of noun phrases within the complementation clauses was analysed based on Rohdenburg's Complexity Principle (1996). Discussion was made in relation to a pattern of linguistic change in progress, densification of content. A diachronic distinctive collexeme analysis (cf. Hilpert, 2006) was performed based on the *-ing* forms of the complementation clauses in order to explore their semantic domain preferences at different time periods. Analysis results were interpreted with respect to the process of language change as they showed that linguistic change in American English is still in progress.

# DEDICATION

To my grandpa in heaven, thank you for being my inspiration today and forever.

# ACKNOWLEDGMENTS

Many thanks to everyone who have encouraged, supported and helped me from the start until the end of my study. The love and care provided each day act as vitamins to keep me going. You remain anonymous but you know who you are!

Special thanks to my supervisors, Dr. Nick Groom and Dr. Neil Millar for their wonderful advices and insightful comments during the reading of my drafts. I would like to express my deepest gratitude for your support and guidance during my entire study. Although I was a slow learner asking many questions but you never gave up on me. You have taught me to become a more independent researcher, to appreciate academic writing and most important, to love corpus linguistics more each and every day.

To my beloved parents, thank you for providing me with financial support while living in the UK for the past six years until today. Your efforts were endless and painful at times but I appreciate every moment of what you have given to me since the day I was born. I promise to work harder in the future to show a better side of myself. Also, thank you for trusting me in my chosen route.

Once again, thank you from the bottom of my heart. I love you all!

# TABLE OF CONTENTS

# FIGURES

# TABLES

# CHAPTER 1

# INTRODUCTION

The aim of this study is to investigate the development of non-finite complementation clauses of two semantically comparable verbs, *prevent* and *stop* based on a large computerised database which consists of four written genres. This study hopes to document interesting instances of frequency change across 19th, 20th and 21st century American English which were not previously noticed in the literature. It is conducted using the framework outlined in Sellgren (2007) which was derived from the work of Mair (1995 and 2002).

English is rapidly evolving over time. This is a fascinating phenomenon because in the long run, certain structural patterns, words or meanings associated with them may have disappeared while new ones have emerged providing surprising results. Hence, the study of certain grammatical patterns is worthwhile and meaningful because language changes over time. According to Aitchison (1991: 4), it would be strange if a language remains unaltered. Bauer (1994: 1) says that English in present day is changing and one can observe the changes that are occurring. Many observation studies of language change have been made especially by Bloomfield (1933), Barber (1964) and Potter (1975). They claim that English has changed since the Middle English period due to the strong influence of media, especially radio and television. Some examples of changes are the use of 'American' *do you have* instead of the established 'British' *have you got* and the downgrading of some full verbs (*get, want, go*) to auxiliary status in some of their uses. Such

changes are massive especially when comparing a variety of English around the world. This is not an unexpected result because language never stops developing.

One area in English grammar which is suspected to be undergoing particularly vigorous change and restructuring is that of non-finite complementation clauses (Leech et al., 2009: 181). Non-finite complementation clauses are clauses without a tensed verb (Carnie, 2007: 204). Algeo (1988: 22) mentions that when exploring grammatical differences between British and American English, the richest area of divergence is verb complementation clauses. This phenomenon was observed through examples of complementation clauses in British English which are less common or non-occurring in American English. Leech et al. (2009: 181) state that contemporary English is characterised by a complex system of non-finite complementation clauses which sets it apart from Old and Middle English and most other European Languages.

In order to identify specific constructions involving non-finite complementation clauses, Leech et al. (2009: 186) suggest some specific superordinate or matrix verbs such as *start, begin, help,* and *prevent*. According to Mair (2006: 119), these non-finite verb forms have become more functionally important especially in discourse since the Middle English period. He states that in spite of the lack of attention that the non-finite verb forms have received in the literature on progress change, there is no indication that the diachronic force which characterised them in the Early Modern English has declined in the recent past. Vosberg (2009: 212) claims that in the process of linguistic change involving non-finite verb forms, major analyses of historical and present-day electronic text corpora show that American English is sometimes lagging behind but

very often it is British English that has been more old-fashioned. He concludes his investigation of non-finite complementation clauses by stating that both British and American English follow the same trends in their development but at different speeds. In contrast, Algeo (2006:1) argues that today's British English is becoming closer to the common familial form of current varieties whereas American English preserves the older uses that have become obsolete in British English.

In Kachru's three model circle of World English(es) (1992: 235), American English lies within the 'Inner Circle' variety and is commonly used as a reference to World English(es). American English creates great impact on influencing other varieties in terms of lexicon, grammar, spelling and pronunciation. The precise reasons for this influence are unclear but reasonable conjectures have been made such as exposure of American English's prestige, spread of popular American culture and global dominance of the American media – Hollywood movies and music industries. According to Graddol (1997: 7), as a consequence of World War II, the United States has become a global economic and cultural presence, thus making American English the dominant world variety.

Previous studies have uncovered interesting tendencies among the *prevent* and *stop* complementation clauses. Mair (1995 and 2002) studies both *prevent* and *stop* by using four corpora which are Lancaster-Oslo-Bergen Corpus (LOB), Freiburg-LOB Corpus of British English (FLOB), Standard Corpus of Present-Day Edited American English for Use with Digital Computers (Brown) and Freiburg-Brown Corpus of American English (Frown). Sellgren (2007) focuses on the complements of *prevent* in 18$^{th}$, 19$^{th}$ and 20$^{th}$ century British English by using the

Corpus of Late Modern English Texts (CLMET). Babováková (2005) investigates the length of complementation clauses between *prevent* and the participle in British and American English by using the International Computer Archive of Modern and Medieval English (ICAME) corpora and the Collins Cobuild corpora. All the results obtained from the studies mentioned show divergent development in British and American English. Mair (2002: 111) comments that these observed regional contrasts are temporary because both varieties are actually developing towards a common goal at different speeds. However, there is still no full account of an integrated synchronic variation and diachronic development of these complementation clauses.

With regards to those previous studies, there is no prior research which focuses on the diachronic development of non-finite complementation clauses in American English across the 19th, 20th and 21st centuries. To fill this gap, I have chosen to investigate their development and concentrate in detail on two semantically comparable verbs, *prevent* and *stop* to uncover any interesting changes which were not previously noted in the literature. This study will employ a corpus-based approach to explore the *prevent* and *stop* complementation clauses when used with and without a preposition or complementiser *from*. This approach serves as a tool for quantitative analysis with the help of a large computerised database, known as corpus. The application of this study is to use the analysis results in structuring language teaching materials and to revise existing grammar books, especially those aimed at non-native English speakers.

The focus of this study is to acquire a diachronic perspective on the trends of the *prevent* and *stop* complementation clauses. In order to obtain empirical evidence, the data will be extracted

from the 400 million word Corpus of Historical American English (COHA) which represents 19th, 20th and 21st century American English. The aim of using COHA is to find out how the frequencies for the *prevent* and *stop* complementation clauses have changed in each decade from 1810 to 2009 and which variant (the with or without *from*) is being preferred in American English. With those data obtained, a test on the length of noun phrases based on Rohdenburg's Complexity Principle (1996) will be performed. This test will observe whether the length of noun phrases increases or decreases when the preposition or complementiser *from* is used. Subsequent follow-up is a diachronic distinctive collexeme analysis based on the *-ing* forms of each complementation clause which is derived from the work of Hilpert (2006). This analysis enables the determination of their semantic domain preferences at different time periods. Lastly, methodological limitations are considered and ways for improvement will be suggested.

A basic assumption in this study is that shifts over time in frequency of the *prevent* and *stop* complementation clauses from COHA indicate linguistic change in progress. Based on this assumption, this study attempts to answer the following research questions:

(i) What frequency changes can be observed across 19th, 20th and 21st century American English for:

- *prevent / stop* + noun phrase + *from* + *-ing* (e.g. She *prevented / stopped it from eating.*)
- *prevent / stop* + noun phrase + *-ing* (e.g. She *prevented / stopped it eating.*)

(ii) Do changes in overall frequency reflect changes in preferences for the with *from* variant and the without *from* variant?

(iii) Is there a connection between the length of noun phrases and the use of preposition or complementiser *from*?

(iv) Do certain *-ing* forms show a preference for the with *from* variant or the without *from* variant and are there changes over time?

This thesis is structured as follows:

Chapter 2 presents an account of the meanings of *prevent* and *stop* from *The New Shorter Oxford English Dictionary on Historical Principles* (1993) and their examples of constructions from various other British and American English dictionaries. The theoretical background for the *prevent* and *stop* complementation clauses is drawn together. My hypothesis for the development of the complementation clauses is included. A separate section is outlined for Rohdenburg's Complexity Principle (1996) which is connected to the length of noun phrases within the complementation clauses and related previous studies. A final section discusses the patterns of linguistic change in progress with particular attention on densification of content as it is linked to the length of noun phrases.

Chapter 3 introduces important terminology such as *clause, complement* and *non-finite* which are used in this study. It then looks into the concepts of animacy and agentivity which are related to the classification of verbs by Biber et al. (1999). This chapter also discusses the foundations of corpus linguistics, reasons for the corpus-based approach and the representativeness of corpus

data. It ends by explaining briefly about COHA which is the main source of data and its representativeness in this study.

Chapter 4 is a short chapter on methodology. It starts by repeating the research questions and provides a schematic overview to illustrate the four distinct methodological stages to guide this study. A set of search queries which were formed is listed and the types of syntactic patterns that were caught are described. Each step on how the data was retrieved from COHA for analysis is specified.

Chapter 5 presents an overall quantitative evaluation of the *prevent* and *stop* complementation clauses. It reports the accelerating or slowing down trends in American English across the 19$^{th}$, 20$^{th}$ and 21$^{st}$ centuries with graphical illustrations. There is also an analysis of the proportional values to compare the with *from* variant with the without *from* variant. This chapter continues with the analysis of the length of noun phrases based on Rohdenburg's Complexity Principle (1996). A discussion on a pattern of linguistic change in progress, densification of content which is related to the analysis of the length of noun phrases is presented.

Chapter 6 continues with the diachronic distinctive collexeme analysis which is derived from the work by Hilpert (2006). This analysis measures the degree of attraction or repulsion of the *-ing* forms to the complementation clauses at different time periods (cf. Stefanowitsch and Gries, 2003). There is an additional discussion of the theoretical background of collostructional analysis

and diachronic distinctive collexeme analysis which is separated from Chapter 2. The following section presents the general methodology of diachronic distinctive collexeme analysis. Then, a related case study which is relevant to this study is discussed. The chapter continues by illustrating the methodology used for the analysis. Finally, it reports and interprets the findings from the analysis.

Chapter 7 concludes this thesis by providing a brief summary of the principal findings and the research significance of this study. It notes down the methodological limitations faced during the study and recommends new directions for future research.

Corpora sources and additional tables from the analyses can be found in the Appendix section.

# CHAPTER 2

# THEORETICAL BACKGROUND

## 2.1 Overview

As indicated in Chapter 1, this study is primarily concerned with non-finite complementation clauses. Specifically, it is an attempt to find out the development of the *prevent* and *stop* complementation clauses in American English across the 19th, 20th and 21st centuries.

This chapter outlines the theoretical background to this study with reference to the relevant literature. The *prevent* complementation clauses are discussed in Section 2.2 and the *stop* complementation clauses in Section 2.3. The meanings for both verbs from *The New Shorter Oxford English Dictionary on Historical Principles* (1993) and their examples of constructions from various other British and American English dictionaries are presented (Sections 2.2.1 and 2.3.1). This is followed by a discussion of previous studies on the development of each complementation clause (Sections 2.2.2 and 2.3.2). My hypothesis for the development of both complementation clauses is included. An overview of Rohdenburg's Complexity Principle (1996) which is connected to the length of noun phrases within the complementation clauses and related previous studies are presented in Section 2.4. Subsequently, a brief discussion on the patterns of linguistic change in progress is presented (Section 2.5) as well as some previous studies of densification of content as it is linked to the length of noun phrases (Section 2.5.1). The final section summarises the chapter (Section 2.6).

## 2.2 *Prevent*

### 2.2.1 Meanings of *prevent* in various dictionaries

*Prevent* is a word that originates from Latin and has entered the English language from the Late Middle English period. There are two different semantic categories within the meaning for *prevent* found in *The New Shorter Oxford English Dictionary on Historical Principles* (1993: 2348) which are related to this study. The first provides the sense of 'to act or do in advance' and the latter gives the meaning of 'to stop, hinder or avoid'. However, both meanings have diffused nowadays due to language change processes and have not shown any clear differences. This diffusion phenomenon is taken into account for this study. Here is an example (*The New Shorter Oxford English Dictionary on Historical Principles*, 1993: 2348):

(1) I shall not *prevent* your going.

In (1), *prevent* gives the sense of stop in advance and avoid going at the same time. Both semantic categories have dispersed and there is much ambiguity to the meaning.

The earliest meaning was first recorded during the Late Middle English period. All related meanings for *prevent* found in *The New Shorter Oxford English Dictionary on Historical Principles* (1993: 2348) are listed in Table 1 below. After every meaning, there are the recorded usage dates. The abbreviations of the recorded usage dates can be found in Appendix 2.

Table 1: Related meanings for *prevent* according to *The New Shorter Oxford English Dictionary on Historical Principles* (1993: 2348)

| **Category** | **Related meanings of prevent** |
|---|---|
| (i)<br>To act or do in advance | 1.<br>(a) Act before, in anticipation of, or in preparation for (a future event, a point in time). LME-E19.<br>(b) Meet beforehand or anticipate (an object, question, desire, etc.). M16-M19.<br>(c) Come, appear, or act before the time or in anticipation. M16-M17.<br>2. Act before or more quickly than (another); anticipate in action. LME-E19.<br>3.<br>(a) Come, arrive, or appear before; precede; outrun, outstrip. L15-M18.<br>(b) Come in front of, meet in front. M16-M17.<br>(c) Outdo, excel. M16-M17.<br>4. Hasten or bring about prematurely; anticipate. M16-M17.<br>5. Occupy or use beforehand. Also, preoccupy or prejudice (in mind). M16- M18.<br>6. Of God or his grace: go before with spiritual guidance or help, in anticipation of human action or need. M16. |
| (ii)<br>To stop, hinder, avoid | 7.<br>(a) Provide beforehand against the occurrence of (something); make impracticable or impossible by anticipatory action; stop from happening. M16.<br>(b) Use of preventative measures. E17.<br>8. Forestall or thwart by previous or precautionary measures. LME.<br>9. Preclude from or deprive of a purpose, expectation. M16.<br>10. Frustrate, defeat, make void (an expectation, plan, etc.). M16-M18.<br>11. Stop (something) from happening to oneself; escape or evade by timely action. L16-E18.<br>12. Cause to be unable to do or be something, stop (followed by *from doing, from being*). Also (with ellipsis of *from*), stop from *doing* or *being.* M17. |

In the *Webster's New Collegiate Dictionary* (1973: 912) which is commonly used in America, *prevent* has several meanings and the one that is related to this study is 'to hold or keep back: hinder / stop'. It states that *prevent* is often used with *from* and continued by a transitive verb. Later, an updated *Webster's New World Dictionary of American English* (1988: 1067) describes the meaning as 'to stop or keep *from* doing something' and 'to hinder'. *The American Heritage Dictionary of the English Language* (1992: 1436) presents an example for *prevent*:

(2) ...*prevented* us *from* winning...

*A Dictionary of American – English Usage: Based on Fowler's Modern English Usage* (1957: 447) explains about the constructions of *prevent*: *prevent* him *from* going or *prevent* his going but not *prevent* him going. The *Longman Modern English Dictionary* (1976: 886) also gives the meaning as 'to cause not to do something' and illustrates it with an example as follows:

(3) Illness *prevented* him *from* going.

The *Longman Dictionary of Contemporary English* (1978: 866) has a category of *prevent from* (verb preposition) and provides an example as shown below:

(4) You can't *prevent* me *from* going there.

In *A Dictionary of English Collocations: Based on the Brown Corpus* (1994: 1463), *prevent* is separated into two categories, *prevent* and *prevented*. Under p*revent*, there is *prevent it from* and *prevent the government from* where both phrases have an equal exclusive and inclusive frequencies of occurring twice in the corpus. Under *prevented*, there is *prevented from, prevented me from* and *be prevented from*. These three phrases have a range of exclusive and inclusive

frequency values of occurring between two to five times in the corpus. There are inputs on *prevent it* and *to prevent it* but I will not comment on them because there are no further explanations to what follows after the pronoun *it*.

In summary, *prevent* means 'to act in advance' or 'to stop / hinder'. From the observations in both British and American English dictionaries, there are no examples of constructions for *prevent* being used without the preposition or complementiser *from.* All the examples mentioned above illustrate that *prevent* is always being used with *from* followed by the *-ing* forms. In order to find out more about when *prevent* favours the variant without *from*, I will look into previous studies which have employed corpus-based approach in order to focus on the grammatical aspects of the *prevent* complementation clauses.

### 2.2.2 Development of *prevent*

According to Quirk et al. (1985: 1194), negative meaning verbs such as *stop*, *prevent* and *prohibit* have a related ditransitive construction in where the preposition *from* precedes the *-ing* forms clause as a second object. For this reason, *from* is optional. An example is provided as follows:

(5) They tried to *prevent the plane from landing* on the runaway. (Quirk et al., 1985: 1194)

Rosenbaum (1967: 89-91) argues that the noun phrase in *prevent*-type verb constructions is termed as a subject because it replaces a pronoun. Below is an example (Rosenbaum, 1967: 90):

(6) Base structure:

I *prevented* [NP [N it] [S [NP John] [VP go]]]

↓

Complementiser Placement

↓

I *prevented* [NP [N it] [S from [NP John] [VP ing go]]]

↓

Extraposition

↓

I *prevented* [NP [N it]] [S from [NP John] [VP ing go]]

↓

Auxiliary Transformation

↓

I *prevented* [NP [N it]] [S from [NP John] [go + ing]]

↓

Pronoun Replacement

↓

I *prevented* [NP John] [S from [VP go + ing]]

↓

Surface structure:

I *prevented John from going*.

In (6), the subject is raised to the matrix clause and this is termed Pronoun Replacement. The subject is raised to the direct object position and *from* acts as a complementiser.

Aarts (1990: 149) states that *prevent*-type verbs are dyadic predicates because they act as a subject or a direct object. *Dyadic predicates* are predicates which take two arguments: a subject argument and a direct object argument (Aarts, 1990: 149). Below are examples (7-9) illustrated by Aarts (1990: 149):

(7) I *prevented the accident*.

(8) I *prevented Andrew's leaving*.

In (7) and (8), *prevent* assigns a direct object role to the noun phrases, *the accident* and *Andrew's leaving*. Sentence (9) below is more problematic.

(9) They *prevented language theory from influencing* the students.

↓

Passivation Process

↓

They *prevented* the students *from being influenced* by *language theory*.

When (9) undergoes passivation process, *prevent* does not assign the direct object role to the noun phrase, *language theory* nor *the students* but rather to the whole string of noun phrase. This means that the noun phrase, *language theory* is a subject of a subordinate clause introduced by the complementiser *from*.

(10) They *prevented language theory Ø influencing* the students. (Aarts, 1990: 149)

When *from* is omitted in active constructions like in (10), the sentence is still grammatically correct and acceptable although (9) and (10) are syntactically different. As a result, *from* cannot be regarded as an inflectional element because there is no evidence in English syntax that *from* can function in that way.

This analysis leads Aarts (1990: 152) to claim that *from* is optional in *prevent*-type verb constructions as *from* is a complementiser. Below are the reasons to support his claim with examples provided (Aarts, 1990: 153):

Firstly, the element of *from* is meaningless – this is a characteristic of most complementisers. There is no way that a clause introducer like *that* carries meaning. A complementiser acts as a clause introducer; it does not carry any importance but signals the presence of a subordinate clause.

Secondly, with active subordinate clauses, *from* in *prevent*-type verb constructions can be omitted as shown in (10). This is similar for other complementiser like *that* which can also be left out as shown in (11).

(11) I believe *that* she is marvellous.

↓

I believe Ø she is marvellous.

Thirdly, it is observed that there is a close connection between the complementisers and the type of clauses they introduce. *From* always takes a non-finite -*ing* clause as shown in (12), but not a finite clause as in (13), nor a non-finite to-infinitival clause as in (14).

(12) I *prevented Kate from eating* the biscuits.

(13) *I *prevented Kate from ate* the biscuits.

(14) *I *prevented Kate from to eat* the biscuits.

Based on the basis of examples above, Rosenbaum (1967: 90) also concludes that '*from* is not an instance of preposition but a complementiser of the basic form '*from -ing*''.

Aarts (1990: 158) continues the analysis by discussing animate and inanimate noun phrases when used with *prevent* under passive condition. The passive of *prevent*-constructions is rather problematic. Rita Manzini in Aarts (1990: 158) suggests two sub-categorisation frames which Aarts follows:

(15) [ - , complementiser phrase (CP)] – inanimate noun phrase

[ - , noun phrase (NP) complementiser phrase (CP)] – animate noun phrase

The examples below explain the suggested sub-categorisation frames (Aarts, 1990: 158):

(16) Andrew was *prevented from leaving* the house.

(17) [-] was *prevented* [$_{NP}$ Andrew$_1$] [$_{CP}$ *from* PRO$_1$ *leaving* the house].

(18) Andrew$_1$ was *prevented* [[$_{spec\text{-}of\text{-}IP}$ t$_1$] *from* t$_1$ *leaving* the house].

↑_____|↑_____|

The noun phrase, *Andrew* in (16) is animate; hence it does not have any problem undergoing a fronted passivation process. As shown in (17), the PRO [1] (null pronoun) subject in the subordinate clause is co-indexed with *Andrew* in the matrix clause. This is called a control structure; it does not involve any phrase movement unlike in (18). The noun phrase, *Andrew* in

[1] PRO is a phonetically null element posited in Principles and Parameters Theory in constructions where a non-finite verb has no overt subject (*Oxford Concise Dictionary of Linguistics*, 2007: 320).

(18) is not suppose to move from the Specifier-of-IP[2] position in the lower clause to the Specifier-of-IP position of that same lower clause and then on to the Specifier-of-IP position of the matrix clause. If the noun phrase has such movement like in (18), it will create an informal development of the grammar which is impossible.

(19) I *prevented* anyone *from leaving*.

↓

*Anyone was *prevented from leaving* by me.

As demonstrated in (19) by Postal (1974: 159-160) in Aarts (1990: 160), not all animate noun phrases can be front passivised. An animate noun phrase which has a non-assertive form (a form that does not assert truth-values for the sentences in which it occurs) cannot undergo this process.

(20) ?*[$_{NP}$ Language theory]$_1$ was *prevented* [$_{CP}$ t$_1$ *from* t$_1$ *influencing* the students].

(21) *[$_{NP}$ The cake]$_1$ was *prevented* [$_{CP}$ t$_1$ *from* t$_1$ *being* eaten].

In (20) and (21) as illustrated by Aarts (1990: 158), inanimate noun phrases of the complementiser clauses cannot undergo front passivation process. The subject of noun phrase of the complement clause cannot be front passivised as this would involve the forbidden movement as shown in (18).

According to Aarts (1990: 164), his analysis shows that *prevent*-type verbs are verbs that are structurally ambiguous. He concludes that the element *from* is best regarded as a complementiser. When *from* is used in *prevent*-type verb constructions, it is optional in active complementation clauses but obligatory in passive complementation clauses. An inanimate noun phrase following

[2] IP is a phrase which is seen as headed by a unit that is typically an auxiliary or inflection of a verb (central to X-bar syntax) (*Oxford Concise Dictionary of Linguistics*, 2007: 203).

*prevent* cannot be fronted under passive condition whereas an animate noun phrase can be fronted.

Huddleston and Pullum (2002: 657) mention that verbs expressing prevention and abstention, such as *prevent*, exists with two variants: *prevent NP from -ing* and *prevent NP -ing*. Many of these verbs permit a direct object which represents the preventee (*prevent him from seeing it*) and the action prevented (*prevent it*). However, not all verbs work like that: *They dissuaded him from doing it*, but not **They dissuaded his action*. Some of them allow a gerund-participial clausal complement without *from* (*prevent the boy leaving*). The function of *from* is as a role of 'intended actions [which act] as spatial goals', with an explanation of 'to hold someone back from doing something is to hold them back so that they will not perform the action'.

The usage of *prevent* with or without *from* started in the Late Modern English period resulting in regional contrasts between British and American English (Mair, 2002: 111). This phenomenon has caused many curiosities (cf. Van Ek, 1966; Dixon, 1995; Rohdenburg, 1995; Mair, 2002; Babováková, 2005; Heyvaert et al., 2005 and Sellgren, 2009) for gerunds dependent such as *prevent*, where there is a variable use of the preposition or complementiser *from*. Sellgren (2009: 14) claims that the diversion between British and American English did not start until late 19th century or early 20th century. British English is increasingly favouring the without *from* variant but this process is unlikely to reach completion, considering the strong preference for the with *from* variant in passivised and complex sentences which are cognitively difficult to process. On the other hand, American English is using the with *from* variant only – American English is

perhaps stricter than British English in its prescriptive guidelines in language use or more standardising in nature.

Mair (2002: 112) investigates this phenomenon of divergence by using the LOB, FLOB, Brown and Frown corpora.

Table 2: Ratio of *prevent NP from -ing* vs *prevent NP -ing* in the LOB, FLOB, Brown and Frown Corpora: frequencies in instances (Mair, 2002: 112)

| **Corpora sampling years** | **British English** | **American English** |
|---|---|---|
| 1961 | 34:7 | 47:0 |
| 1991/92 | 24:24 | 36:1 |

Based on Table 2, in British English, it is observed that *prevent NP from -ing* shows a decrease (from 34 to 24 instances) while *prevent NP -ing* (from 7 to 24 instances) increases considerably. In American English, the trend still favours *prevent NP from -ing* (47 against 0 instance) in comparison to *prevent NP -ing* (36 against 1 instance). Mair then compares his results with two major corpora of spoken English and finds that the results are rather close. In the London-Lund Corpus of Spoken English which has half a million words, there are 6 instances of *prevent NP from -ing* as against 5 instances of *prevent NP -ing*. In the Corpus of Spoken Professional American English (2+ million words), there are 35 instances of *prevent NP from -ing* but none of *prevent NP -ing*. Hence, Mair (2002: 115) concludes that in the future, British English may continue to have two variants while American English will maintain its prepositional variant as dominant as both variants are diverging.

According to Dixon (1995: 217), American English may be the only variety that favours complementation clauses with *from -in*g among the various varieties of English in the world. The without *from* variant occurs in present-day Australian English apart from British English. Rohdenburg (2009: 211) suggests that the negative feature in verbs of leaving (eg. *depart*, *resign* and *escape*) which is associated with *from* may play a special role in the complex American English constructions as it is supported by some parallel findings in the area of sentential complementation. He points out that verbs of negative causation such as *prevent*, *stop* and *save* have followed a trend in American English where they mark negative orientation by the use of *from* without exception whereas in British English, the verbs of negative causation use *from* only in the passive. In contrast, the *Oxford English Dictionary* (2010, online version) claims that *prevent NP -ing* is actually a shorter or reduced form of *prevent NP from -ing*.

Rudanko (2006: 45) studies the *-ing* forms complements in British and American English and compares the pace of change between British and American English. He argues that the difference between both varieties of English was noticeable as early as the nineteenth century but the change in American English has been completed more fully at present time than in British English. American English may have led the way in the emergence of the *-ing* forms due to regional variation. This was supported by the results from his study which show an extraction of *to -ing* complements instead of *to infinitive* complements in 19$^{th}$ century American data. This occurrence was due to both syntactic and semantic factors such as metaphorical movement. However, he suggests that larger-scale corpora should be used for further investigation to observe any later development in contemporary American English.

To summarise, *prevent* must always occur with *from* in passive complement clauses which are cognitively harder to process (Sellgren, 2009: 14). American English strongly favours the use of the with *from* variant. This is due to the negative causation verbs like *prevent* which have followed a trend in American English where the use of *from* plays a special role in those negative constructions. Nevertheless, there are still no studies which investigate the development of the *prevent* complementation clauses diachronically in American English. This gap needs to be filled. My study aims to find out how the *prevent* complementation clauses develop in American English across the 19th, 20th and 21st centuries. My hypothesis is *prevent NP -ing* will decrease in comparison to *prevent NP from -ing* due to the obligatory status of *from* in American English. *Prevent NP from -ing* is expected to occur more frequently in American English especially towards the 21st century as suggested by Mair (2002) and Sellgren (2009). As the hypothesis remains unconfirmed, this study will seek evidence in support of it.

## 2.3 *Stop*

### 2.3.1 Meanings of *stop* in various dictionaries

*Stop* has Germanic origins and started its usage from the Late Middle English period. There are many meanings for *stop* found in *The New Shorter Oxford English Dictionary on Historical Principles* (1993: 3074) as it is a multi-purpose word[3]. However, only those meanings which are relevant to this study are listed. Below is an example from *The New Shorter Oxford English Dictionary on Historical Principles* (1993: 3074):

[3]Multi-purpose word is a word which serves several or many purposes in terms of its meanings (*The New Shorter Oxford English Dictionary on Historical Principles*, 1993: 1855).

(22) You have *stopped* me from brooding.

In (22), *stop* gives the sense of block in advance which is related to *prevent* in Section 2.2.1. There are also other meanings which are irrelevant as they are semantically different such as the following (*The New Shorter Oxford English Dictionary on Historical Principles*, 1993: 3074):

(23) He brought the car to a jolting *stop*.

(24) Come and *stop* with us in September.

In (23), *stop* denotes a halt in a journey while in (24) *stop* indicates staying in as a visitor. Both meanings are not related to this study. Thus, they are not included.

The earliest meaning of *stop* and its usage which emphasizes the preposition or complementiser *from* was first recorded in the Late Middle English period. Table 3 below shows the related meanings for *stop* according to *The New Shorter Oxford English Dictionary on Historical Principles* (1993: 3074). Every meaning that is listed in Table 3 has its recorded usage dates too. The abbreviations of the recorded usage dates can be found in Appendix 2.

Table 3: Related meanings for *stop* according to *The New Shorter Oxford English Dictionary on Historical Principles* (1993: 3074)

| **Related meanings of *stop*** |
|---|
| 1. (a) Prevent or stem the passage of; dam or block the channel of (running water etc.); keep out or exclude (light, the weather, etc.). LME.<br>(b) Staunch (bleeding, blood). L16. |
| 2. (a) Check or impede the onward movement of; bring to a standstill or state of rest; cause to halt on a journey. Also, prevent the departure of. LME.<br>(b) Check (an opponent, a stroke, a blow, etc.) with a counter movement or stroke; counter (a blow etc.). E18.<br>(c) Shoot or bring down (game, a bird). Also, (of a bullet or wound) arrest the rush of (a charging enemy or wild animal) with rifle-fire. M19. |
| 3. (a) Withhold or deduct (a sum of money) in paying wages or repaying a debt, in order to cover rent, special clothing, etc. LME.<br>(b) Withhold (goods) as security or in lieu of payment. M18. |
| 4. (a) Cause (a person) to desist from or pause in a course of action or conduct. Frequently followed by *from*, *in*. LME.<br>(b) Hold (a thing) in check; cause (a thing) to cease action. LME.<br>(c) Cause (a person) to break off from speaking or pause in a conversation. M16. |
| 5. (a) Restrain or prevent (a person) from an intended action. Also followed by *from*. LME.<br>(b) Stay or suspend (proceedings); prevent (a decree etc.) from taking effect. L17.<br>(c) Of a camera: give a still picture of (a moving object). M20. |
| 6. Hamper or impede the course or progress of (affairs, a project, etc.); hinder (a person). LME-E18. |
| 7. (a) Put an end to (a movement, activity, course of events). LME.<br>(b) Prevent the onset of. M16. |
| 8. (a) Discontinue (an action, a sequence of actions, work, etc.). E16.<br>(b) Put an end to the issue or supply of (an allowance etc.). M19. |
| 9. Cause (a machine or piece of mechanism) to cease operation. M16. |
| 10. Intercept and detain (goods, post, etc.) in transit. E17. |

As mentioned in 2.2.1, the *Webster's New Collegiate Dictionary* (1973) is a common dictionary used in America. In *Webster's New Collegiate Dictionary* (1973: 1446), *stop* has several meanings such as 'to hinder or prevent the passage, to hold back – restrain / prevent, to cease an activity or operation – discontinue, to come to an end – pause / hesitate / halt'. An updated *Webster's New World Dictionary of American English* (1988: 1321) provides almost the same meanings for *stop*. Surprisingly, there are no given examples of constructions which state that *stop* must be used together with the preposition or complementiser *from*. *The American Heritage Dictionary of the English Language* (1992: 1771) presents some examples for *stop*:

(25) ...tried to *stop* the bleeding...

(26) ...*stop* running...

(27) ...*stopped* the check...

The examples provided once again, did not mention about *stop* being used with *from*. Hence, this causes much confusion.

On the other hand, the *Longman Modern English Dictionary* (1976: 1092) provides the meaning for *stop* as 'to prevent from moving or acting' and illustrates it with an example:

(28) He won't *stop* me *from* going.

One of the meanings for *stop* in the *Longman Dictionary of Contemporary English* (1978: 1098) is 'to prevent' but it is demonstrated with an example without the preposition or complementiser *from* as follows:

(29) You must *stop* her telling them.

On the contrary, there is a special section which discusses its usage. It mentions that there are differences in meanings for diverse *stop* constructions. One illustration is as follows:

(30) He *stopped* me *(from)* listening. (*Longman Dictionary of Contemporary English*, 1978: 1098)

Example (30) has an equivalent meaning of 'he didn't allow me to listen'. The preposition or complementiser *from* is however bracketed which gives an impression of being optional.

In *A Dictionary of English Collocation: Based on the Brown Corpus* (1994: 1698-1699), *stop* is divided into three categories: *stop, stopped* and *stopping*. Under *stop*, there is *to stop bleeding* which has equal exclusive and inclusive frequencies of occurring twice in the corpus. Under *stopped*, *stopped pacing* and *stopped beating* have low frequencies too. Under *stopping*, *was stopping* has a frequency of two exclusively and inclusively. There is no record of *stop* being used with *from*.

In summary, *stop* means 'to hinder / prevent / discontinue'. Based on the observation in various British and American English dictionaries, there are fewer examples of constructions for *stop* being used with *from* unlike *prevent NP from -ing*. This means that the question to when *stop* will be followed by *from* and when will not, remains unanswered. For this reason, I want to find out when *stop* is being used with the preposition or complementiser *from* like *prevent* and the development of the without *from* variant. I will use some previous studies to guide me through.

### 2.3.2 Development of *stop*

*Stop* is a semantically complex verb. According to Huddleston and Pullum (2002: 657), *stop* is listed together with *prevent* as verbs expressing prevention and abstention which is related to the similar ditransitive construction. They state that the use of *from* is developed via the moderately transparent metaphor of intended actions as spatial goals. Although many of the preventive verbs allow only a direct object signifying the preventee or the action prevented, some preventive verbs also allow a gerund-participial clausal complement without *from*: *stop them doing it.* This summarises that *stop* can be followed by *from* or without *from* in complementation clauses.

Huddleston and Pullum (2002: 1238) claim that *stop* is a problematic member of the *prevent*-type verbs and does not always allow a genitive in its construction. Below are some examples (Huddleston and Pullum, 2002: 1238) to illustrate that *prevent* cannot be replaced by *stop*:

(31) He *prevented us from seeing her*.

↓

Replacement

↓

He *stopped us from seeing her*.

(32) He *prevented our seeing her*.

↓

Replacement

↓

*He *stopped our seeing her*.

In (31), *us* is an object pronoun followed by *from* whereas in (32), *our* is a possessive pronoun which is not followed by *from*. Hence, when *stop* replaces *prevent*, the meanings change completely and it is impossible to replace in (32).

*Stop* has a wide range of meaning as illustrated by Huddleston and Pullum (2002: 1238) below:

(33) We must *stop him coming back tomorrow*.

(34) They *stopped us playing* before we had finished the first set.

*From* can be inserted in (33) because *stop* in (33) means 'not allow and prevent'. In (34), *from* cannot be inserted as it belongs to the causative of an aspectual verb such as *keep* and *start*. *Stop* in (34) means 'made us stop'. Both (33) and (34) demonstrate that *stop* has different meanings. Quirk et al. (1985: 1194) also classify *stop* under aspectual verbs where it is followed by a complementation of *-ing* participle clause. Huddleston and Pullum (2005: 295) define *aspectual verb* as a verbal category mainly to indicate the speaker's view of the temporal structure on how a clause describes a situation such as habitual or complete. The English progressive (*-ing* forms) is an example of an aspectual verb conveying a situation as being in progress.

Mair (2002: 112-113) investigates *stop* in the same terms as he investigates the grammatical phenomenon of *prevent*, as discussed in Section 2.2.2. He mentions that several other synonyms of *prevent* such as *stop* or *block* are increasingly attested in the without *from* variant in British English nowadays but does not give any comment for American English. However, in his findings, only *stop* occurs frequently enough to provide a complete set of data for his matching corpora. The corpora that were used are LOB, FLOB, Brown and Frown.

Table 4: Distribution of *stop NP -ing* in the LOB, FLOB, Brown and Frown corpora: frequencies in instances (Mair, 2002: 113)

| **Corpora sampling years** | **British English** | **American English** |
|---|---|---|
| 1961 | 4 | 0 |
| 1991/92 | 17 | 0 |

Table 4 shows his results for *stop NP -ing* from four corpora which demonstrate an increase (from 4 to 17 instances) in British English. Nothing is shown for American English (0 instances); this phenomenon needs more attention. If as assumed by Mair (2002: 113), the core verbs, *prevent* and *stop* are leading a grammatical change prior to lexical diffusion[4], they should move towards the without *from* variant in British English. Further similar verbs such as *deter, dissuade, discourage, free, hinder* and *prohibit* which originate from the Latin and Germanic language families are expected to show the same patterns. These verbs are not treated in his findings because they are too rare in the corpora.

Mair (2006: 132) once again studies *stop*. He compares *stop NP from -ing* with *stop NP -ing*.

Table 5: Ratio of *stop NP from -ing* vs *stop NP -ing* in the LOB, FLOB, Brown and Frown Corpora: frequencies in instances (Mair, 2006: 132)

| **Corpora sampling years** | **British English** | **American English** |
|---|---|---|
| 1961 | 6:4 | 5:0 |
| 1991/92 | 3:12 | 7:0 |

Table 5 above demonstrates the results which Mair has obtained. The results support his assumption in his previous study (2002). *Stop NP -ing* has emerged as a syntactic Briticism in

[4] Lexical diffusion is a phenomenon of gradual spread of phonological changes across the vocabulary of a language (Brinton and Traugott, 2005: 12).

the past half-century (from 4 to 12 instances) in contrast to *prevent NP -ing* (refer discussion in Section 2.2.2). As this variant does not occur even once in the Brown and Frown corpora, American English perhaps prefers the traditional *stop NP from -ing* variant (from 5 to 7 instances). These results exhibit a clear-cut difference between British and American English grammatical constructions. In a historical study of constructions using the Oxford English Dictionary (OED) Baseline 1900 corpora (Mair, 2006: 133), *stop NP from -ing* provides ambiguous results as there are not enough data. *Stop NP -ing* and *stop NP's -ing* are observed to occur only once each.

Historically, the spread of *from* complements is thought to have started in American English and this spread was characterised as less clear in British English (Mair, 2002: 113). The phenomenon of *from* being a competitor in both varieties of English is interesting. Sellgren (2010: 49) suggests examining the semantic distinction between both competing complementation clauses. Her suggestion is inspired by Bolinger's (1968) generalisation that a difference in syntactic form entails a difference in meaning. She proposes that *from -ing* and Ø *-ing* may be developing different meanings as both variants are diverging. She links the *from -ing* variant to the idea of hypotheticality which is a condition less likely to be met, indicated by the *-ing* form. On the contrary, the Ø *-ing* variant is linked to the idea of a realised event or an existing property of the object noun phrase, indicated by the object noun phrase and the *-ing* form together. Her study shows that a large proportion of the data appeared to follow this semantic distinction. She summarises that the distribution of both variants may not be stable and possibly they may select their own functions of use, semantically or otherwise determined.

Dixon (1991: 236) also provides an explanation to support the semantic distinction generalisation between *from -ing* and Ø *-ing*. When *from -ing* is used, the agent denoted by the subject in the sentence employs indirect means. Rudanko (2002: 58) comments that *from -ing* may involve indeterminacy. In contrast, when Ø *-ing* is used, the agent employs direct means and this lies within sentences with underspecified contexts. He adds that Ø *-ing* has a sense of immediacy and external observability. The semantic distinction of *from -ing* and Ø *-ing* can apply to only animate and non-abstract object noun phrases or 'preventees' as the notion of direct or indirect indicates an act of prevention. Hence, this distinction can only apply to certain complementation clauses. In addition, Dixon (1995: 217) mentions that the omission of *from* after *prevent, stop, save* and *spare* can occur in British and Australian English, but not in American English.

Unsurprisingly, this particular contrast between British and American English remains a mystery. A frequently used American English dictionary, the *Webster's Dictionary of English Usage* (1989: 770) in Mair (2002: 115) claims that the without *from* variant is quite uncommon in American English but in recent British usage it is a familiar trend. Leech et al. (2009: 197) argue that the without *from* variant is an emerging candidate of rapid grammaticalization. The term *grammaticalization* is defined as a part of the study of language change that is concerned with how lexical items and constructions come in certain linguistic contexts to serve grammatical functions and how grammatical items develop new grammatical functions (Hopper and Traugott, 2003: 1). Leech et al. (2009: 197) find that there is an increase in usage of *stop -ing* (from 24 to 43 instances) in American English based on Brown and Frown corpora. They provide an example that is commonly used in present-day English but do not clarify which variety of English it is. The example is as follows:

(35) We *stopped them walking* up the street. (Leech et al., 2009: 197)

The rapid spread of the without *from* variant in British English that is unparallel to American English at present state, distinguishes this variation from the eighteenth-century variation. This means that British English is standardising on one variant and American English on the other. Based on the findings, they predict that this dynamic trend will continue to have two options in British English while American English will only have one option in the twentieth century.

In summary, *stop* has a wider range of usage where *from* can be omitted as it is categorised under aspectual verbs. In Mair's findings, however, there is no occurrence of *stop NP -ing* for American English. Rudanko claims that *stop NP -ing* gives a sense of directness and immediacy to the agent while *stop NP from -ing* shows uncertainty. The question of how the *stop* complementation clauses have developed over time in American English remains unanswered. For this reason, I would like to investigate the trends of both variants diachronically in order to observe their development. Therefore, my hypothesis for *stop NP from -ing* is that it will gradually occur more often in American English. *Stop NP -ing* will not be favoured in American English even towards the 21st century. American English will keep the traditional variation of using the with *from* variant and it is expected to increase in frequency tremendously over time because this pattern is not strongly influenced by theory of grammaticalization. As the hypothesis needs further investigation, this study will seek evidence in support of it.

## 2.4 Complexity Principle

The length of noun phrases within the non-finite complementation clauses is connected to Rohdenburg's Complexity Principle (1996). This is defined as:

> In the case of more or less explicit grammatical options the more explicit one(s) will tend to be favoured in cognitively more complex environments. (Rohdenburg, 1996: 151)

In simpler words, the more explicit variant is preferred in more complex environments and has a tendency to be treated as more formal than the less explicit one. Examples of complex environments are passivised sentences, relativisations, extractions and sentences with long and complex object noun phrases. The more complex an object noun phrase is, the longer it takes to recognise the relationship and to identify the function of the verb. That complex object noun phrase eventually becomes harder to process and requires more time. To simplify this process, the complex object noun phrase will choose the more explicit variant to be understood easily.

The principle of Complexity Principle is inspired by Hawkins' research (cf. 1990 and 1992). Although differences in grammatical explicitness may be achieved in many ways, Rohdenburg (1996: 151) suggests one simple way which is to count the number of words in each construction and to compare the variants which are related to each other. He states that the use of the Complexity Principle is to distinguish clearly between the more explicit variant and the less explicit one. The more explicit variant is normally represented by the bulkier element or construction. However, this principle faces some problems such as word order variants and grammatical signals. Word order may generally be assumed to be logically equivalent. Below are examples by Rohdenburg (1996: 150):

(36) She put the fire *out*.

(37) She put *out* the fire.

In (36), the particle *out* is associated with *put* follows by the object, *fire*. In (37), the two elements, *put* and *out* form a continuous construction. Grammatical signals may be deleted or added as shown by the following examples (Rohdenburg, 1996: 151):

(38) I helped him *to* write the paper

(39) I helped him write the paper.

In (38) and (39), the difference is an optional grammatical signal, the infinitive marker *to*. This principle may also conflict with other factors, in particular with stylistic and semantic tendencies. Despite the problems faced, the principle is still useful in many applications.

Rohdenburg's Complexity Principle (1996) is applicable in the *prevent* and *stop* complementation clauses. Sellgren (2010: 49) explains that this principle predicts that the more explicit variant which is *from -ing* can be found more often in complex syntactical environments such as long noun phrases. This is because the more explicit variant will be easier to process cognitively in complex structures. They decrease the processing load created by the structural complexity. To illustrate, Sellgren (2010: 50) provides an example from the BNC:

(40) But the fight did not *prevent* the fundamental beliefs in the nation and 'the historic integrity of the island of Ireland', as nationalist parties described it in their New Ireland Forum (1983-1984: i.28), *from remaining* basic to the perceptions of both parties. (BNC, A07, 317)

In (40), the object noun phrase of *prevent* is very long but *from -ing* completes the understanding of the message clearly. Visser (1973: 2372) also claims that *prevent* always selects *from -ing* for long object noun phrases as this variant started spreading diachronically since the 17th century.

In a study done by Rohdenburg (2006), he emphasises that in the case of more or less explicit variants, the more explicit one will tend to be preferred in cognitively more complex environments. He then provides some examples for *stop* (Rohdenburg, 2006: 63):

(41) They *stopped all the passengers from continuing* their journey.

(42) They *stopped all the passengers Ø continuing* their journey.

(43) *All the passengers were stopped from continuing* their journey.

In all the examples above, *stop* belongs to a sense of negative causation. These examples suggest that both British and American English have shown a strong tendency, between the 18th and 20th centuries, to replace the less explicit option – without *from* (42) by the more explicit option – with *from* (41), with the cognitively more complex passive construction (43) leading the development. However, the examples do not in themselves indicate neither diachronic change nor regional similarity or variation.

In general, the passive constructions in American English like (43) have developed completely in comparison to the active constructions especially the without *from* variant like (42). Rohdenburg (2006: 64) analyses British English's active and passive constructions by using two traditional tabloids, *The Daily Mail* and *The Mail on Sunday* (1993).

Table 6: Comparison of passive and active constructions with *stop NP from -ing* and *stop NP -ing*: frequencies in instances (Rohdenburg, 2006: 64)

| **Construction types** | **with *from*** | **without *from*** | **Total** | **Percentage of *from*** |
|---|---|---|---|---|
| Passive | 16 | 0 | 16 | 100% |
| Active | 131 | 840 | 971 | 13.5% |

The results from Table 6 show that in present-day British English, the without *from* active variant is used increasingly based on the number of samples collected (840 instances). On the other hand, passive constructions prefer the with *from* variant over the without *from* variant (16 against 0 instances).

According to Rohdenburg (2006: 52), there are three degrees of complexity in noun phrases involving subject pronouns. Below are examples of the use of subject pronouns in main clauses by Rohdenburg (2006: 52):

(44) *My mates (they)* made a collection for me.

(45) *My mates at work (they)* made a collection for me.

(46) *The fella that done the damage (he)* said to me...

The examples distinguish between the different degrees of complexity – simple noun phrase (44), complex noun phrase (45) which involves an additional preposition together with noun phrases, and very complex noun phrase (46) that contains relative clauses. On the basis of these examples, he separates the noun phrases from the active constructions in Table 6 above and compares the average number of words within those noun phrases in Table 7 below.

Table 7: Average number of words of noun phrases in the *stop* active constructions: frequencies in words (Rohdenburg, 2006: 64)

| **Average number of words** | ***from*** | **without *from*** |
|---|---|---|
| Simple noun phrases | 2.24 | 1.54 |
| Full noun phrases (excluding personal pronouns) | 2.81 | 2 |

It is observed that the with *from* variant occurs with longer objects. The average number of words for full noun phrases is higher (2.81 words) than for simple noun phrases (2.24 words). The results from the comparison of active and passive constructions and the average number of words for noun phrases support his theory of Complexity Principle.

Babováková (2005) investigates the length of noun phrases within *prevent*-type constructions with participle *-ing* in British and American English based on Rohdenburg's Complexity Principle (1996). She finds that noun phrases with more than four items and above tend to favour the more explicit variant, *from -ing*. However, when the noun phrase consists of only one word, this more explicit variant, *from -ing* still dominates. This means that theoretically, short object noun phrases could also support the use of *Ø -ing* which represents the less complex environment. Roughly about 57% to 70% of the data found were used with *from* regardless of long or short object noun phrases. It is only in passivisation that the Complexity Principle is firm. The mystery to what factors actually promote the use of *from* still remains unanswered.

With regards to Rohdenburg's Complexity Principle (1996), working memory is related in the processing of the more or less explicit grammatical options. The term, working memory was introduced by Miller, Galanter and Pribram (1960: 16) to support their concept of 'plans' for the

performance of sequential operations. The original definition of *working memory* was a special state or place where a plan is remembered while being executed (Miller, Galanter and Pribram, 1960: 65). Kim et al. (2011: 2) define *working memory* as a limited capacity processing system that supports the online maintenance and manipulation of information over short time periods. In simpler terms, working memory is the ability to hold information in attention and to process it. Sometimes, working memory is also known as short-term memory.

In an investigation of sentences containing long distance dependencies by King and Just (1991), they claim that a language system utilizes a general, non-specialized working memory device for information storage and processing. During the interaction of these processes, complex cognitive phenomena arise. King and Just examine the processing of centre-embedded relative clauses with different working memory capacity. Examples of the clauses (King and Just, 1991: 581) are as follows:

(47) The reporter [that *attacked* the senator] admitted the error.

(48) The reporter [that the senator *attacked*] admitted the error.

(47) is a subject relative clause where the subject, *the reporter* functions together with the relative clause verb *attacked* in the embedded clause. (48) is an object relative clause where the subject, *the reporter* is the direct object of the relative clause verb *attacked* in the embedded clause. They find that the event-related brain potentials components are larger for words in object relative clauses like in (48). This suggests that these sentences place greater processing demands on which working memory than subject relative clauses like in (47). When sentence patterns are more complex and difficult to process, greater working memory is being used as an

essential tool to overcome the problem because more cognitive demands are needed. The query to how working memory undergoes the processing of the Complexity Principle has not been researched much.

As there are still many unanswered questions regarding the Complexity Principle and how working memory processes the complexity of the noun phrases, I would like to investigate and hope to answer some of them. As claimed in the studies mentioned, for both *prevent* and *stop* complementation clauses, the more explicit option is *from -ing* and the less explicit option is Ø -*ing*. Based on Rohdenburg's Complexity Principle (1996), I will study the length of noun phrases within the complementation clauses. I will observe how the length increases or decreases when used with and without the preposition or complementiser *from*. My hypothesis is as the length of noun phrases increases, the preference to use *from* in the complementation clauses will become stronger. Hence, I will explore further based on the analysis in Chapter 5.

## 2.5 Patterns of linguistic change in progress

As mentioned in Chapter 1, the basic assumption in this study is that shifts over time in frequency of the *prevent* and *stop* complementation clauses from COHA indicate linguistic change in progress. According to Croft (2000: 1), the process of language change is a historical phenomenon. A *language* is defined as a system of rules and forms divided into phonology, morphology, syntax, semantics and pragmatics which is treated as an entity (Croft, 2000: 2). A *change* is defined as an innovation or an act of individual speakers that has been widely adopted by members of a community (Janda and Joseph, 2003: 13). The process of language change

happens via the reproduction of entities which are the utterances produced in the context and the speakers' grammatical knowledge of the language found inside their minds. Paris (1868) in Janda and Joseph (2003: 7) criticises that 'the development of language does not have its causes in language itself but rather in the physiological and psychological generalizations of human nature'. Based on this critique, Bonfante (1946: 295) expresses that 'languages are historical creations'.

As the grammar of Standard English keeps changing, Svartvik and Leech (2006: 206) suggest that linguistic changes are related closely to the process of language change. These changes tend to pursue some particular patterns which are listed below:

(i) Grammaticalization – A famous process of language change where items of vocabulary are gradually getting subsumed into grammatical forms.

(ii) Densification of content – The process of compacting meaning into a smaller number of words.

(iii) Colloquialization – The use of written grammar tends to become more colloquial or informal like speech.

(iv) Americanization – The use of grammar in World English especially British English tends to follow the Americans' trend.

Svartvik and Leech (2006: 211) observe that in the present-day grammatical change, English is becoming a more democratic language because it appears to lack the honorific forms which indicate the relationship of superiority or inferiority, deference or familiarity, between speakers.

It uses more colloquialised lexis such as semi-modal verbs and contraction forms. It is also turning into a non-sexist language by using more gender-neutral nouns and pronouns in its vocabulary (Svartvik and Leech, 2006: 215). There is no English Academy to regulate the language but it is the public opinions which bring about a shift in the usage of vocabulary and grammar.

There have been many corpus-based studies on noun phrases which are related to one of the suggested patterns, densification of content – Johansson (1980) on plural attributive nouns, Biber and Clark (2002) and Biber (2003) on comprehension and complexity in noun phrases as well as Leech et al. (2009) on compactness of noun phrase structures. As I will be exploring the length of noun phrases in the *prevent* and *stop* complementation clauses, I would like to find out the reasons as why they increase or decrease in length. Do they follow some particular patterns like densification of content? This question remains unanswered and needs further investigation. The next section discusses some previous studies which are related to densification of content.

### 2.5.1 Densification of content

According to Leech et al. (2009: 249), *densification* is defined as a process of compacting meaning into a smaller number of words as a countervailing influence in the noun phrase, alongside colloquialization. In this study, I will refer to densification as *densification of content* which is more appropriate as the process is based on the context of the constructions. As mentioned in Section 2.5, there have been many corpus-based studies on noun phrases which are related to densification of content.

Leech et al. (2009: 249) suggest that the process of densification of content does not limit itself only to noun phrases analysis but a more general trend – the measurement of lexical density on the corpora. *Lexical density* is defined as an increase in the proportion of content words (information content carriers) to function words (grammatical function carriers), which in turn indicates a tendency to condense more information into a smaller number of words (Leech et al., 2009: 208). The process of measuring lexical density can be achieved by calculating the number of lexical word tokens (mostly nouns, verbs, adjectives and adverbs) as a proportion of the total number of word tokens in a corpus. Leech et al. (2009: 249) define *lexical words* as any words tagged as adjective, noun, lexical verb, adverb, number, letter of the alphabet or other formulaic expression.

Biber et al. (1999: 65) study lexical density through lexical words by using the Longman Spoken and Written English Corpus (LSWE) which contains approximately twenty million words of text from four different genres. Amongst the genres (conversation, fiction, newspaper language and academic prose) in LSWE, they find that nouns are the most frequent lexical word class with an average of one in every four words. Nouns are most common in newspaper language (54%) and least common in conversation (35%). In newspaper language and academic prose, there are three to four nouns per lexical verb which go along with adjectives (modifier of nouns). This lexical density phenomenon focuses on the transmission of information where nouns are fundamental and important in this environment. The findings demonstrate that a high percentage of nouns corresponds to longer clauses, embedded clauses and complex phrases. In contrast, the low frequency of nouns in conversation shows lower density of information because conversations are linked closely to lexical verbs where speech meanings are expressed through verbs.

In English, the written prose has experienced extensive stylistic change over the past centuries (Biber, 2003: 169). Written prose in the 17th century differed from conversational prose and their grammatical features became more distinct from speech starting in the 18th century. Starting from the 18th and 19th centuries, there was an increase use of elaborated noun phrases, passive verbs and relative clause constructions in the newspaper language. However, towards the 20th century, there is a mark of change where the newspaper language starts to drift towards more speech-like styles – colloquialization (cf. Biber and Finegan, 1997/2001). Nevertheless, the highly dense use of nouns and integrated noun phrase constructions remain generally in newspaper language.

Biber and Clark (2002) argue that from a historical perspective, newspaper language has consistently used noun-noun sequences to a great extent. Their argument is supported by Biber's findings (2003: 174-177) based on A Representative Corpus of Historical English Registers (ARCHER) which covers a range of samples from both British and American English. The results show a huge increase of the use of noun-noun sequences starting from 1900. Noun phrase structures are used extensively in newspapers with both pre- and post-modifiers when compared with three other genres (drama, fiction and medical). Below is an example by Biber (2003: 174):

| | | | |
|---|---|---|---|
| (49) | *pre-modifiers* | *noun* | *post-modifiers* |
| | A Quaker-run training | college | for teachers |

Biber finds that about 40% of all pre-modifiers nouns come from the newspapers and they tend to be longer than other genres. These pre-modifiers are used to compressed information into relatively few words. In (49), the information can be expanded into *The college specializes in*

*training and that is run by Quakers*. In addition, many of these nouns are productive in combining with other different head nouns as shown below (Biber, 2003: 174):

(50) government + agencies, control, decision

(51) business + administration, cards, ideas

(52) TV + ads, channel, crew

Post-modifiers again are also being compressed into few words to pack extra information about the respective nouns. Most of the time, they appear as embedded noun phrases instead of separate clauses.

Biber (2003: 170) claims that the preferred linguistic style of newspaper discourse is influenced by two competing forces – popularization and economy. *Popularization* is often associated linguistically with colloquilization and this term is defined as using the kind of language that is most accessible to a mass readership (Leech et al., 2009: 218). *Economy* means condensing information into a small linguistic area – another kind of pressure that affects newspaper language (mostly in headlines and running texts) (Leech et al., 2009: 218). There is a high demand from readers for newspaper language to communicate information efficiently and economically and thus, resulting in compressed styles which depend heavily on noun phrase constructions. Nowadays, newspaper language opts to exhibit an intermediate style between both forces which is to express verbally in some aspects and at the same time, to be literate in other aspects. In modern day newspaper language, newspapers have been very innovative in the development of noun phrase constructions when compressing expressions.

The results from Biber et al. (1999) and Biber (2003) are not surprising as Svartvik and Leech (2006: 209) also find that in newspaper reporting – especially in headlines – it is common for a sequence of several nouns without any intervening words to occur. More information is packed together in newspaper headlines through the compression of nouns in noun phrases for more efficient and economical communication. In newspapers, each column is limited in space and thus it is useful to compact nouns together in order to save the space for other information. Svartvik and Leech (2006: 209) present a couple of examples such as the following:

(53) New York City Ballet School instructor

(54) Real estate tax shelter sales people

(55) San Francisco Redevelopment Agency Executive Director Chuck Springfield

The long noun phrase in (53) can be spread out to become more explicit by adding in prepositions and articles – *An* instructor *at the* School *of* Ballet *in the* City *of* New York. This is similar for (54) and (55) where both of the compact noun phrases can be extended. These high frequencies of noun-noun sequences have also increased substantially in other types of information-oriented written languages and not only in newspaper reporting.

As discussed above, previous findings exhibit evidence that noun phrases are used commonly in present-day writings and they follow a particular pattern which is densification of content. This pattern is related to the process of language change. In many accounts, densification of content moves in opposite direction with colloquialization. Both of these factors need further investigation based on more recent corpora. I, however, believe that both factors will create

productive awareness among speakers of English. Under any circumstances, I would like to discuss my analysis of the length of noun phrases in relation to densification of content.

## 2.6 Summary

In this chapter, first the meanings and examples of constructions of *prevent* and *stop* were presented from various British and American English dictionaries. *Prevent* is a less problematic verb in comparison to *stop* as all the examples given do not use the preposition or complementiser *from*. *Stop*, on the other hand, has a mixed of examples being used with and without *from*. Then, related previous studies were discussed. Rohdenburg's Complexity Principle (1996) was reviewed together with previous investigation in relation to the length of noun phrases within the complementation clauses. They were linked to the patterns of linguistic change in progress. One of the patterns, densification of content was found to be closely related.

Based on the discussion, my hypothesis for both *prevent* and *stop* complementation clauses is the 19th, 20th and 21st century American English will favour the with *from* variant over the without *from* variant. *Prevent / stop NP from -ing* will increase whereas *prevent / stop NP -ing* will decrease. This may be due to the trend in American English that for verbs of negative causation, they mark negative orientation by the use of *from* without exception (Rohdenburg, 2009: 211). The *prevent* complementation clauses are expected to have steadier development over the centuries whereas the *stop* complementation clauses will still be developing. This hypothesis will be tested in Chapter 5.

The next chapter introduces some useful terminology and discusses Biber et al.'s classification of verbs (1999) which will be used for the diachronic distinctive collexeme analysis in Chapter 6. It also briefs about the foundations of corpus linguistics, representativeness of corpus data and Corpus of Historical American English (COHA) as the main source of data.

# CHAPTER 3

# TERMINOLOGY AND DATA

## 3.1 Overview

This study is situated within the field of language change; that my hypothesis is changes in the trends of the overall frequency or semantic domain preferences indicate linguistic change in progress. Denison (1998: 93) mentions that it is useful to track frequencies of occurrence from Early Modern English through to present day. Leech et al. (2009: 9) add that large machine-readable textual data will help the systematic compilation of statistics and frequencies. Corpora will provide aids in refining the understanding of recent and ongoing grammatical change in Standard English.

This chapter starts by defining the important terminology which is used throughout the study (Section 3.2). A discussion on the concepts of animacy and agentivity which are related to Biber et al.'s classification of verbs (1999) is presented (Section 3.3). The next section (3.4) explains the foundations of corpus linguistics and reasons to employ the corpus-based approach and continues with the representativeness of corpus data (Section 3.4.1). Then, it also discusses the 400 million word Corpus of Historical American English (COHA) which is the main source of data (Section 3.5) and describes its representativeness in this study (Section 3.5.1). The last section summarises the chapter (Section 3.6).

## 3.2 Terminology

Important terminology which is used throughout this study is defined and explained to provide a clearer picture – *clause, complement* and *non-finite*.

Carnie (2007: 201) defines a *clause* as a syntactic unit whose structure includes a subject (usually a determiner phrase that has the property indicated by the predicate) and a predicate phrase (a group of words that assign a property to the subject). An example of a clause is a simple sentence.

(56) The boy ran. (Carnie, 2007: 201)

In (56), *the boy* is the subject and *ran* is the predicate phrase. A clause which stands on its own is known as a *main clause* (root or matrix clause) (Carnie, 2007: 202). (56) is an example of a main clause. Sometimes, there are clauses hidden within clauses. An example is as follows:

(57) [Peter said [that Danny danced]]. (Carnie, 2007: 202)

In (57), there are two clauses. *Peter said that Danny danced* is the main clause. *(that) Danny danced* is another clause which sits inside the main clause. The hidden clause is known as an *embedded clause* (subordinate clause) (Carnie, 2007: 202). *Danny* is the subject and *danced* is the predicate phrase.

Embedded clauses are distinguished among complement, adjunct and specifier clauses (Carnie, 2007: 203). A *complement clause* is a clause which acts as a sister to the verb, also known as a

complement. It functions as an argument of a matrix verb. It can occupy the subject argument slot as in (58) or the object slot as in (59). Below are examples from Kidd et al. (2005: 50):

(58) [That Rufus was late] angered his boss.

(59) Rufus could see [that he has angered his boss].

In (58), *that Rufus was late* is a complement clause because *that* is a complementiser, *Rufus* is a subject and *was late* is a predicate phrase. In (59), *that* as a complementiser is optional. An *adjunct clause* is a clause which appears in an adjunct position. A relative clause is an example of an adjunct clause.

(60) [The man [I saw get into the cab]] robbed the bank. (Carnie, 2007: 203)

In (60), *I saw get into the cab* which is a relative clause, modifies the head *the man.* A *specifier clause* is a clause which serves as the subject of a sentence. An example is as follows:

(61) [[For Mary to love that boor] is a travesty]. (Carnie, 2007: 203)

In (61), *for Mary to love that boor* is a specifier because it serves as the subject of the sentence.

In short, there are two basic types of clauses – main and embedded. Embedded clauses lie within main clauses. There are three types of embedded clauses – complement, adjunct and specifier. Table 8 summarises the types of clauses.

Table 8: Summary of clause types (Carnie, 2007: 203)

| **Types of clauses** | | | |
|---|---|---|---|
| Main | Embedded | | |
| | specifier | complement | adjunct |

Complementation clauses can be identified within different syntactic distinctions – finite and non-finite (Kidd et al., 2005: 50). A *finite complementation clause* is a complement clause which has a predicate that is tensed. A *non-finite complementation clause* is a complement clause without a tensed verb. Below are some examples from Kidd et al. (2005: 50):

(62) I see [Elena walked to the institute this morning].

(63) Mike enjoys [chatting with his students].

In (62), the complementation clause is finite because the subordinate verb is tensed. In (63), the complementation clause is a non-finite clause; subordinate verb is tenseless.

Finite and non-finite complementation clauses are not subject to the same grammatical constraints. One main difference is non-finite complementation clauses frequently lack an overt subject and do not require subject-verb agreement. There are some tests obtained from Radford (1988) in Carnie (2007: 204) that are used to distinguish between finite from non-finite complementation clauses. One way is to look for the agreement and tense morphology on the verb. When the tense is changed to the past, a difference emerges. Examples from Carnie (2007: 204) are as follows:

(64) I know [you *eat* asparagus]. → I know you *ate* asparagus.

(65) I've never seen [you *eat* asparagus]. → *I've never seen you *ate* asparagus.

Finite complementation clause allows past tense morphology change from *eat* to *ate* but non-finite complementation clause do not allow it. In (64), *eat* can be changed to *ate* but not in (65). Hence, the complement clause in (64) is finite whereas in (65), it is non-finite.

English verb inflections are mainly divided in two categories – tensed / finite and tenseless / non-finite (Huddleston, 1984: 386). Tensed verbs are basically divided into two types – past tense and present tense. Present indicative has two forms – 3rd person singular and general (everything else). Present subjunctive and imperative use similar forms as present indicative general because there are no verbs in English which are distinct. Hence, there is no ground for making an inflectional difference. Tenseless verbs are divided into three forms – base (infinitive), *-ing* (gerund and present participle – no English verbs are again distinct in these two forms) and *-en* (past pasticiple). Table 9 presents a summary of verb inflections by using the verb *take*.

Table 9: Summary of verb inflections by using the verb *take* (Huddleston, 1984: 286)

| **Verb categories** | | | ***take*** |
|---|---|---|---|
| Tensed / finite | Past tense | | *took* |
| | Present tense | 3rd person singular | *takes* |
| | | General | *take* |
| Non-tensed / non-finite | Base form | | *take* |
| | *-ing* form | | *taking* |
| | *-en* form | | *taken* |

## 3.3 Animacy and agentivity

Previous studies like Bybee et al. (1994) suggest that the progressive forms were originally limited to animate and agentive subjects. They underwent grammaticalization process during the Modern English period and started occurring with inanimate and non-agentive subjects (Hopper and Traugott, 2003: 104). Hundt (2004: 47) considers this aspect of progressive forms as one of the most researched categories in grammaticalization theory. The *progressive* form is defined as a subject [which] is located in the midst of doing something (Bybee et al., 1994: 136). This form

provides a location for the subject and activity in a construction. Thus, it contains either explicitly or implicitly the following elements of meanings:

(i) an agent

(ii) is located spatially

(iii) in the midst of

(iv) an activity

(v) at reference time.

According to Hundt (2004: 47), the roots of the modern progressive form started from the Middle English period or perhaps from the Old English period. However, the rules to use the aspectual form emerged only in the 17$^{th}$ century. After the aspectual function of the progressive form grammaticalised, its usage increases in frequency. Based on the evidence from the Early Modern English part of the Helsinki corpus, Elsness (1994: 11) in Hundt (2004: 48) finds that the progressive form is used three times as frequent in 1640-1710 as in 1570-1640.

Rosenbach (2002: 42) defines *animacy* as an inherent property of concepts which can be assigned to noun phrases and it is a semantic category. Nouns like *fire* and *river* are categorised as [-animate] while *people* is always categorised as [+animate] regardless of the syntactic context. Hundt (2004: 49) claims that agentivity is 'not a property that can be assigned to noun phrases out of context'. These features of agentivity are assigned in the context of a sentence and associated with a semantic role of a constituent. Van Oosten (1984: 326) in Hundt (2004: 49) comments that the responsibility of agentivity is higher than animacy; it is the central

characteristic of the classical 'Agent role' that has the primary responsibility for the occurrence of the action of the predicate. Volition, control and responsibility are properties of classical Agents but it can also be argued that these three characteristics are animate Agents which involve animacy. However, Comrie (1989: 62) sees animacy as a natural property of noun phrases and argues that a high degree of animacy is necessary for a noun phrase to be interpreted as having a high degree of control. An example is provided as follows:

(66) He was *killed* by *lighting*. (Comrie, 1989: 62)

In (66), the noun *lighting* is inanimate. It is responsible for causing the death and has controlled this process. In simpler words, the notion of control, volition and animacy vary independently of one another.

Yamamoto (1999: 149) claims that agentivity presupposes animacy. He also mentions about cases where physically inanimate noun phrases which function as agents always involve figurative language usage or the 'metaphorical touch of inferred animacy' (Yamamoto, 1999: 153). An example is presented as follows:

(67) A lilac bush before the window was moving in the breeze, and *the shadows of its twigs* were *netting* the sun-beams on the wall as they *danced* to and fro. (ARCHER,1880bell.f7)

In (67), the noun phrase *the shadows of its twigs* is an Agent-like role as the verb *netting* requires some sort of Agent. This semantic role is strengthened by the coordination with the verb *danced* which naturally requires an Agent as a subject. According to *Collins Cobuild English Language*

*Dictionary* (1987: 354), the definition of *dance* implies a human Agent – 'when you dance, you move your body and feet in a way which follows a rhythm, usually in time to music'. From this example, it is understood that both animacy and agentivity are associated closely.

As the progressive becomes more grammaticalised, Strang's study (1982) shows that it increases over time and reaches the 'maturity level' of the construction around the middle of the 19th century. Based on the assumption that the progressive was initially used to express activities, she made a hypothesis:

> The construction should be largely confined to use with subjects which are either human or otherwise viewed as capable of agentivity (quasi-human or animal). (Strang, 1982: 443)

Her results verify the hypothesis that inanimate subjects appear only during the end of the 18th century. Denison (1998: 150) points out that an increase of progressive with inanimate or non-agentive subjects may have supported the spread of passive marking. Below is an example:

(68) The house is *being* built. (Denison, 1998: 150)

The passive construction (68) has a non-agentive subject. Example like (68) begins to carry a greater risk of ambiguity for normal progressive to occur with non-agentive subject. This unstable phenomenon raises questions to when exactly the progressive with inanimate and non-agentive subjects starts to increase and how does it affect the non-finite complementation clauses. It needs further investigation to fill the gap.

Biber et al. (1999: 360-364) introduces a classification of verbs which comprises of seven major semantic domains. They domains are as follows:

(i) Activity – verbs that denote actions and events which are associated with choice and take a subject with the semantic role of agent (eg. *run, draw, take*)

(ii) Communication – verbs that involve communication activities (eg. *speak, advise, talk*)

(iii) Mental – verbs that denote activities and states experienced by humans which do not involve physical actions but includes cognitive and emotional meanings expressing various desires (eg. *know, hear, regret*)

(iv) Causation – verbs that indicate a new state of affairs brought by humans or inanimate entity (eg. *enable, allow, force*)

(v) Occurrence – verbs that report events of non-volitional activity and often their subject has the semantic affected role (eg. *happen, become, develop*)

(vi) Existence and relationship – verbs that report a state which exists between entities (eg. *be, seem, appear*)

(vii) Aspectual – verbs that characterize the stage of a process of an activity (eg. *begin, continue, stop*)

The classification of verbs is based on their core meanings which is the meaning that the speakers tend to think of first. There are some verbs in which are not classified according to those domains because they are inappropriate. Nevertheless, there is no single correct classification as their core meanings can belong to more than one category. Some verbs also have secondary meanings.

Based on the query about how progressive forms affect the complementation clauses, I will use the *-ing* forms from the diachronic distinctive collexeme analysis to classify them into different semantic domains according to Biber et al.'s classification of verbs (1999). This methodology will answer my curiosity of the following questions: (1) do certain *-ing* forms show a preference for the with *from* variant or without *from* variant and (2) are there changes over time. The classification of verbs into different semantic domains is suitable for this analysis because the *-ing* forms are based on the core meanings in their sentences. These core meanings correspond to the most typical usage of the *-ing* forms and belong to only one semantic domain (Biber et al., 1999: 361). The discussion of the diachronic distinctive collexeme analysis theoretical background, related previous studies, methodology, case study and analysis results is separated in Chapter 6 to avoid confusion and to obtain clear-cut results from the analyses of the distribution of frequencies.

## 3.4 Foundations of corpus linguistics

Before moving on, it may be useful to provide a brief review of the historical background of corpus linguistics, probably as it relates to the study of diachronic change. In the 1950s, linguists like Harris, Firth and Hill were highly influenced by the behaviourist views of science and looked upon a 'corpus' as the main explanation in linguistic studies. Harris (1957) defines categories in terms of distributional evidence while Firth (1957) investigates the meaning of a word by examining its collocations based on corpora.

On the contrary, Chomsky (1962: 159) in Leech (1991: 8) has different opinions where he criticises the inadequacy of corpus linguistics and put forward the adequacy of intuition. He mentions that the natural corpus data are distorted due to some sentences that will not occur in a corpus because they are understandable, some are false and others are impolite. His criticisms are quite true because at that time, there are only the 'shoebox corpora' available where linguists will fill paper slips in shoeboxes rather than computer as a mean of data storage. The sizes are relatively small and they are used for the study of distinguishing features in phonetics (Lin, 1999: 240). Nonetheless, linguists like Jesperson (1909-1949) and Fries (1952) in McEnery (2006: 3) use the paper-based corpora to study grammar.

Sinclair (1996) defines a *corpus* as a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language. Between 1961 and 1964, Nelson Francis and Henry Kučera first created the Brown Corpus, named after University of Brown (USA) for digital computer usage. Subsequently, the Brown Family of Corpora which contains British English and American English data was built based on the original model of the Brown Corpus. In modern linguistics, a *corpus* is defined as a collection of machine-readable authentic texts including transcripts of spoken data which is sampled to be a representative of a particular language or language variety (McEnery et al., 2006: 5).

Corpora are always associated with linguistics. Hoffmann et al. (2008: 18) define *corpus linguistics* as the systematic study of linguistic phenomena using (machine-readable) collections

of authentic language use such as corpora. In this 21st century, there are many types of corpora available such as general corpora, specialized corpora, bilingual corpora, dialect corpora and learner corpora. These electronic corpora have many advantages - more reliable and accurate results can be produced at minimal cost with the rapid speed of data processing. They are also timesaving in compilation of data. As a result, the usage of corpora is increasing especially for pedagogy, lexicography, translation, forensics linguistics and language variation studies due to the development of computer technology and internet.

There are different approaches to produce results in any study of linguistic phenomena. In my study of linguistic change in progress based on non-finite complementation clauses, I have employed the corpus-based approach as to follow up the works of Sellgren (2007) and Mair (1995 and 2002). Tognini-Bonelli (2001: 74) defines *corpus-based approach* as to expound on, test or exemplify existing theories and descriptions that were not necessarily derived with initial reference to a corpus. In simpler term, it means to use a corpus as a resource to seek evidences of pre-corpus descriptions of language. This approach yields reliable quantitative analysis based on statistical trends, usually with a tagged corpus (at least for part of speech (POS)) and information regarding grammatical feature frequency. An advantage of using the corpus-based approach is that one can find differences in the corpus which intuition cannot perceive (cf. Francis, Hunston and Manning, 1996). Biber et al. (1998: 3) mention that the corpus-based approach provides a means of handling large amounts of language and keeping track of many contextual factors simultaneously. Linguists such as Mair (2006) and Leech et al. (2009) have used this corpus-based approach to investigate diachronic language change while Biber et al. (1999) studied language variation. One disadvantage which can be argued is this approach discourages new

insights as the scope of findings is to some extent predetermined by preconceptions held by researchers.

Below are the general principles of the methodology in corpus linguistics:

First, selected samples from written or spoken texts are compiled to form a corpus. Then, observations of linguistic behaviour represented by the corpus form the linguistic findings. These findings are recorded. They allow linguists to draw conclusions about language variation and language change in general.

Based on these principles, corpus linguistics is an essential quantitative and qualitative methodology in linguistics for the analysis of linguistic features because it provides an empirical foundation to many linguistic researches rather than focusing solely on the theories. In this study, I have chosen to employ the corpus-based approach for accurate and precise observation and analysis of data. It will be based on the 400 million word Corpus of Historical American English (COHA) which was created by Mark Davies.

### 3.4.1 Representativeness of corpus data

What does representativeness mean in corpus linguistics? Biber (1993: 243) refers to *representativeness* as the extent to which a sample includes the full range of variability in a population. A *corpus* is a sample of a language or language variety. Sampling is involved during

the compilation of any corpus of a living language. A corpus must at least have some of the aspects listed below in determining representativeness:

(i) The boundaries of the population – what texts are included and excluded from the population

(ii) Hierarchical organization within the population – what text categories are included in the population and what are their definitions

(iii) Range of linguistic distributions in the population – what different linguistic features are distributed within texts, across texts, across text types

Leech (1991: 27) adds that a corpus is representative when all the findings which are based on the analysis can be generalised in the target language as a whole or specific part of it. Claims of corpus representativeness should be interpreted in relative terms and considered as a statement of faith rather than as fact as there is no objective way to measure its representativeness at present day (McEnery et al., 2006: 21).

According to McEnery et al. (2006: 13), in most corpora, representativeness is determined by two factors: the range of genres included in a corpus (balance) and how the texts for each genre are selected (sampling). The balance of a corpus is determined by its intended usage and usually covers a wide range of text categories which are supposed to be representatives of a language variety. A general corpus which contains both written and spoken texts is normally considered as balance. Controlling the balance of a corpus must be undertaken as an important criterion in designing a corpus and cannot be left out. A legitimate goal should be set before and during the process of corpus compilation in order to be a representative of a language variety. Sampling

depends on the size of the corpus. How large the corpus is depends on the number of texts and the extent the texts should be sampled – full texts or in chunks. To produce a balanced corpus, the proportion of samples and number of texts in each category for sampling needs are important.

The representativeness of a corpus relies on the range of linguistic features in the population. Biber (1993: 243) mentions that during corpus designing, the variability can be considered from situational or linguistic perspectives and they are vital in determining representativeness. There are two corresponding criteria:

(i) External – situationally distribution of linguistic features (eg. genres, registers)

(ii) Internal – linguistically distribution of linguistic features (eg. text types)

Otlogetswe (2004) in McEnery et al. (2006: 14) argues that the internal criteria can be used as a measure of corpus representativeness. Word distributions in a corpus are either distorted to certain varieties or levelled together to represent the corpus entirely. This reflects the stability of the corpus design, whether the usage is to represent the corpus as a whole or to focus on some particular genres only. However, McEnery et al. (2006: 14) view the internal criteria as problematic because if the distribution of linguistic features is predetermined, it is pointless in analysing that corpus to discover naturally occurring linguistic feature distributions. That corpus has been skewed by design. Sinclair (1995) in McEnery et al. (2006: 14) mentions that the texts should be selected according to the external criteria in order for their linguistic characteristics to gain independence during the selection process. Biber (1993: 256) sums up by stating that 'the compilation of a representative corpus should proceed in a cyclical fashion'.

The term, representativeness also covers the types of corpora – general and specialized corpora. McEnery et al. (2006: 15) defines *general corpora* as a basis for an overall description of a language variety. It proportionally should cover as many text types as possible in order for the corpus to be representative maximally of that particular language variety which it is supposed to represent. Formality, informality and different ages / gender / origin of the authors have to be taken into account too. A well-known general corpus is the British National Corpus (BNC) – it is compiled of written texts and transcripts / speech in modern British English to represent as wide range of modern British English as possible. *Specialized corpora* tend to be domain or genre specific and cover specific text types which represent a particular language variety only (McEnery et al., 2006: 15). Some examples of specialized corpora are the HKUST Computer Science Corpus (1 million word corpus of written English texts in the field of computer science, cf. Fang, 1992), the Corpus of Professional Spoken American English (1 million word each of two main sub-corpora which is constructed from a selection of transcripts of interactions in professional settings, cf. Barlow, 1998) and the Michigan Corpus of Academic Spoken English (1.7 million words of contemporary university speech within the domain of the University of Michigan, cf. MICASE Manual, 2002).

The appropriateness of corpus size depends greatly on the purpose of the investigation. The first generation corpora in the early 1960s had only 1 million words, partly because the storage capacity of the computer was quite limited. By the mid-1980s, with the increase in storage capacity, the corpus size increases too. Even so whether the corpus size is 400 million words or 1 million words, it should depend on the frequency and distribution of the linguistic features in that particular corpus as well as other practical considerations. Leech (1991: 10-12) observes that size

is not all-important. For example, in order to study a specific linguistic feature like present-tense verbs in English, a small corpus which contains sufficient examples is good enough for its purpose. Shimazumi and Berber-Sardinha (1996) in McEnery et al. (2006: 72) claim that a small specialized corpus serves a very different yet important purpose from a large multimillion words corpus. However, to conduct a lexical study, then a much larger corpus should be used as it will focus more on the frequency of the distribution of lexis and distinguishing with other lexis of the same category (Santos, 1996: 11). McEnery et al. (2006: 73) summarise that the most reasonable size of a corpus is purely determined by the addressing research questions.

Representativeness covers the standardisation on language varieties too. When the Brown corpus was first released to the public, Francis (1965: 273) describes the corpus as 'a standard corpus of edited American English'. However, this term 'standard' was unclear as Francis (one of the architects of the Brown corpus) was actually referring it as 'a common body of material on which studies of various sorts can be based, among which comparisons can be made'. Ironically, this term 'standard' is a rather controversial definition, whether it is specifying English or other languages. Quirk and Stein (1990: 123) define *Standard English* as the kind of English which draws least attention to itself over the widest area and through the widest range of usages. On the other hand, Crowley (1989: 82) says that *Standard English* means more uniform to some linguists while less prone to structural variation for others. It is the variety which is viewed as most prestigious and authoritative. Nevertheless, it is more practical to connect both meanings together for the complete definition of Standard English.

To sum up, a corpus is thought to be 'representative' when all the criteria – range of genres, texts for each genres, range of linguistic features, types of corpora, standardisation on language varieties are put together during the creation of a corpus taking into account the language variety that is supposed to represent. Ultimately, the research questions will play an important role in determining the appropriateness of the corpus to represent the investigation. Each research topic specifies on different areas in linguistics, thus different types of corpora such as general or specialized corpora will be used. Hence, it is vital to decide for a suitable corpus in order to provide the best result before starting an investigation.

## 3.5 Corpus of Historical American English (1810-2009)

As the corpus-based approach has been employed, the primary data is a corpus of written 19$^{th}$, 20$^{th}$ and 21$^{st}$ century American English – Corpus of Historical American English (COHA). COHA is the latest collection of the historical American English samples, compiled by Mark Davies from Brigham Young University, Provo (Utah) with funding from the US National Endowment for the Humanities. COHA is the largest structured corpus of historical English. It is available free on the web. The alpha version was released in 2010 and the final version was released recently.

COHA contains approximately 400 million words with more than 100000 individual texts, ranging from 1810 to 2009. Each decade has a balanced of texts from four different genres (fiction, magazines, newspapers and non-fiction). Each genre contributes to 25% of the whole

corpus. The main sources which are obtained for the different genres are listed as follows (refer http://corpus.byu.edu/coha/):

(i) Fiction – Project Gutenberg (1810-1930), Making of America (1810-1900), scanned books (1930-1990), movie and play scripts, Corpus of Contemporary American English (1990-2010)

(ii) Magazines – Making of America (1810-1900), scanned and PDF (1900-1990), Corpus of Contemporary American English (1990-2010)

(iii) Newspapers – PDF > TXT of a least five newspapers (1850-1980), Corpus of Contemporary American English (1990-2010)

(iv) Non-fiction – Project Gutenberg (1810-1900), www.archive.org (1810-1900), scanned books (1900-1990), Corpus of Contemporary American English (1990-2010)

### 3.5.1 Representativeness of COHA in this study

COHA is chosen as 'representativeness' for this study. The representativeness of COHA is explained in terms of 'Standard American English', corpus type, text types and corpus size.

COHA matches the definition of 'Standard English'. It represents 'Standard American English' which is an important aspect in this study because this study focuses on 19$^{th}$, 20$^{th}$ and 21$^{st}$ century American English. COHA is compiled within selected American English texts. These texts are produced by novelists, authors, academics, journalists and reporters from a range of

different ages and genders. They are highly skilled professionals in their own fields, hence the texts were written with proper descriptive grammar and uniformed references.

COHA is a general corpus because it contains a range of text types produced in America. COHA has samples from a broad range of different genres and domains as the texts range from 1810 to present-day American English. There are four main genres (fiction, magazines, newspapers and non-fiction), each contributing 25% to the entire corpus and is balanced equally per decade. COHA allows me to investigate the complementation clauses within different genres in the corpus from the early 19$^{th}$ century to the 21$^{st}$ century. With the results obtained, a comparison of the differences or similarities within both the *prevent* and *stop* complementation clauses can be made.

In summary, my choice of COHA is suitable for the purpose of this study. As COHA contains various text types from four different genres in 'Standard American English' starting from the 19$^{th}$ century to the early 21$^{st}$ century, it allows me to observe the development of the complementation clauses per decade. The diachronic results may or may not give an indication that linguistic changes are in progress – this indication depends on the analyses results in Chapter 5 and 6. There will not be any restriction during the process of data extraction as COHA is freely available on the web at all time. COHA is considered as a complete corpus which accommodates to my research questions.

## 3.6 Summary

This chapter started by providing the definitions for some important terminology (*clause, complement* and *non-finite*) which were used throughout the study. It continued by discussing animacy and agentivity which were related to the *-ing* forms within the complementation clauses. Biber et al.'s classification of verbs (1999) was introduced as it will be used to classify the *-ing* forms into different semantic domains for the diachronic distinctive collexeme analysis. Next was a discussion on the foundations of corpus linguistics, its general principles and reasons to why I employed the corpus-based approach for this study. The representativeness of corpus data in terms of range of genres, texts for each genre, range of linguistic features, types of corpora, standardisation on language varieties was explained. Lastly, I also gave a brief introduction about COHA and discussed it as representativeness for this study.

The next chapter specifies the search queries which were formed for this study and explains the methodology used to extract the complementation clauses from COHA.

# CHAPTER 4

# METHODOLOGY

## 4.1 Overview

This chapter presents the search queries and methodology which were used for this study. First, the research questions are repeated to guide this study (Section 4.2). The next section (4.3) details all the search queries that were formed and describes the types of syntactic patterns that will be caught by the search queries. This is followed by a description of the steps taken to retrieve the data from the Corpus of Historical American English (COHA) for analysis (Section 4.4). Finally, a summary of the chapter is provided (Section 4.5).

## 4.2 Research questions

As noted in Chapter 1, this study attempts to explore how the *prevent* and *stop* complementation clauses have developed over time in American English by considering the following research questions:

(i) What frequency changes can be observed in 19th, 20th and 21st century American English for:

- *prevent / stop* + noun phrase + *from* + *-ing* (e.g. She *prevented / stopped it from eating.*)

- *prevent / stop* + noun phrase + *-ing* (e.g. She *prevented / stopped it eating.*)

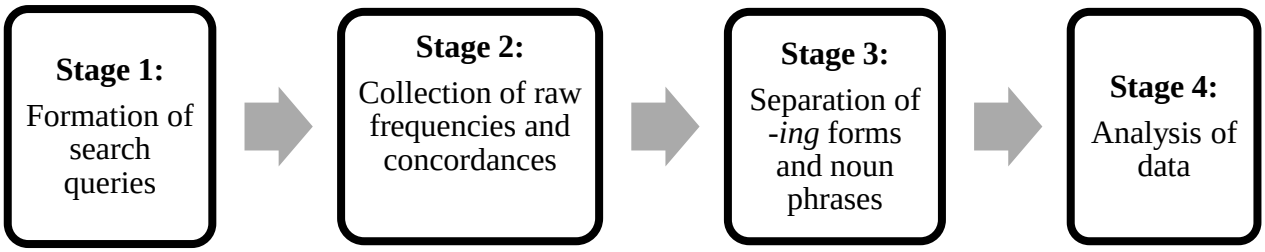
(ii) Do changes in overall frequency reflect changes in preferences for the with *from* variant and the without *from* variant?

(iii) Is there a connection between the length of noun phrases and the use of preposition or complementiser *from*?

(iv) Do certain *-ing* forms show a preference for the with *from* variant or the without *from* variant and are there changes over time?

Based on the research questions, the methodology to collect the data for this study is planned and designed. It is divided into four distinct stages as illustrated in Figure 1 below. These stages will be explained in the following sections.

Figure 1: Schematic overview of four distinct methodological stages



## 4.3 Search queries

First, in order to extract the data from COHA, the search queries were formed. COHA uses the CLAWS7 tagset (Davies, 2009: 164) (refer http://ucrel/lancs.ac.uk/claws/). Table 10 below lists the meanings of the tags and metacharacters that were used in the search queries.

Table 10: Tags and metacharacters used in the search queries

| **Tag** | **Description** |
|---|---|
| [ ] | lemma<br>Example: [prevent] – different inflectional forms of *prevent* such as *prevent, prevents, prevented, preventing* |
| * | zero or more characters<br>Example: [a*] – any types of *article* such as *general article* or *singular article* |
| \| | alternatives<br>Example: [a*]\|[d*] – any *article* or *determiner* |
| p | pronoun |
| a | article |
| d | determiner |
| n | noun |
| jj | adjective |
| m | number |
| r | adverb |
| z | alphabet |
| v*g | -*ing* forms |

Table 11 below provides all the formed search queries and their equivalent meanings. One constraint of the software is that it was not possible to group all of these queries as a single query. Therefore, each query was carried out separately. I have only listed the search queries for *prevent NP from -ing* in the table below. All the other sets of complementation clauses follow the same format except for the replacement of *prevent* to *stop* and omission of the preposition or complementiser *from*. All the queries were set to a maximum of three nouns as standardisation. The maximum use of other grammatical features such as adjectives is two.

Table 11: Search queries used in Corpus of Historical American English (COHA)

| **Search queries** | **Search focus** |
|---|---|
| [prevent] [p*] [from] [v*g] | pronouns |
| [prevent] [n*] [from] [v*g] | single noun |
| [prevent] [n*] [n*] [from] [v*g] | double nouns |
| [prevent] [n*] [n*] [n*] [from] [v*g] | triple nouns |
| [prevent] [a*]\|[d*]\|[jj*]\|[m*]\|[p*]\|[r*]\|[z*] [n*] [from] [v*g] | single (article or determiner or adjective or number or pronoun or adverb or alphabet) + single noun |
| [prevent] [a*]\|[d*]\|[jj*]\|[m*]\|[p*]\|[r*]\|[z*] [n*] [n*] [from] [v*g] | single (article or determiner or adjective or number or pronoun or adverb or alphabet) + double nouns |
| [prevent] [a*]\|[d*]\|[jj*]\|[m*]\|[p*]\|[r*]\|[z*] [n*] [n*] [n*] [from] [v*g] | single (article or determiner or adjective or number or pronoun or adverb or alphabet) + triple nouns |
| [prevent] [jj*] [jj*] [n*] [from] [v*g] | (double adjectives) + single noun |
| [prevent] [jj*] [jj*] [n*] [n*] [from] [v*g] | (double adjectives) + double nouns |
| [prevent] [jj*] [jj*] [n*] [n*] [n*] [from] [v*g] | (double adjectives) + triple nouns |
| [prevent] [a*]\|[d*] [jj*] [n*] [from] [v*g] | (article or determiner + single adjective) + single noun |
| [prevent] [a*]\|[d*] [jj*] [n*] [n*] [from] [v*g] | (article or determiner + single adjective) + double nouns |
| [prevent] [a*]\|[d*] [jj*] [n*] [n*] [n*] [from] [v*g] | (article or determiner + single adjective) + triple nouns |
| [prevent] [a*]\|[d*] [jj*] [jj*] [n*] [from] [v*g] | (article or determiner + double adjectives) + single noun |
| [prevent] [a*]\|[d*] [jj*] [jj*] [n*] [n*] [from] [v*g] | (article or determiner + double adjectives) + double nouns |
| [prevent] [a*]\|[d*] [jj*] [jj*] [n*] [n*] [n*] [from] [v*g] | (article or determiner + double adjectives) + triple nouns |
| [prevent] [a*]\|[d*] [m*] [n*] [from] [v*g] | (article or determiner + number) + single noun |
| [prevent] [a*]\|[d*] [m*] [n*] [n*] [from] [v*g] | (article or determiner + number) + double nouns |
| [prevent] [a*]\|[d*] [m*] [n*] [n*] [n*] [from] [v*g] | (article or determiner + number) + triple nouns |

In this study, embedded clauses are not included in the search. I have formed an example to illustrate a construction with an embedded clause:

(69) He *prevented the girl* who lives next to his house *from dating* his best friend.

The embedded clause in (69) which is *who lives next to his house* will not be caught by the formed search query because the maximum number of nouns that will be catch is three only. The embedded clauses are classed as minority within this study; they therefore have small effect on the results. To quantify 'small effect', I provide an example by using the search query for pronouns ( [prevent][p*][from][v*g] ). There are 17 embedded clauses found within the total amount of 3893 complementation clauses (0.44% of the search). 0.44% is less than1%; hence it does not affect the results much.

To clarify the meaning of the search queries that were formed, I provide an example as below:

(70) [prevent] [a*]|[d*]|[jj*]|[m*]|[p*]|[r*]|[z*] [n*] [from] [v*g]

This search query (70) will catch the meaning of the following:

Find within the corpus,

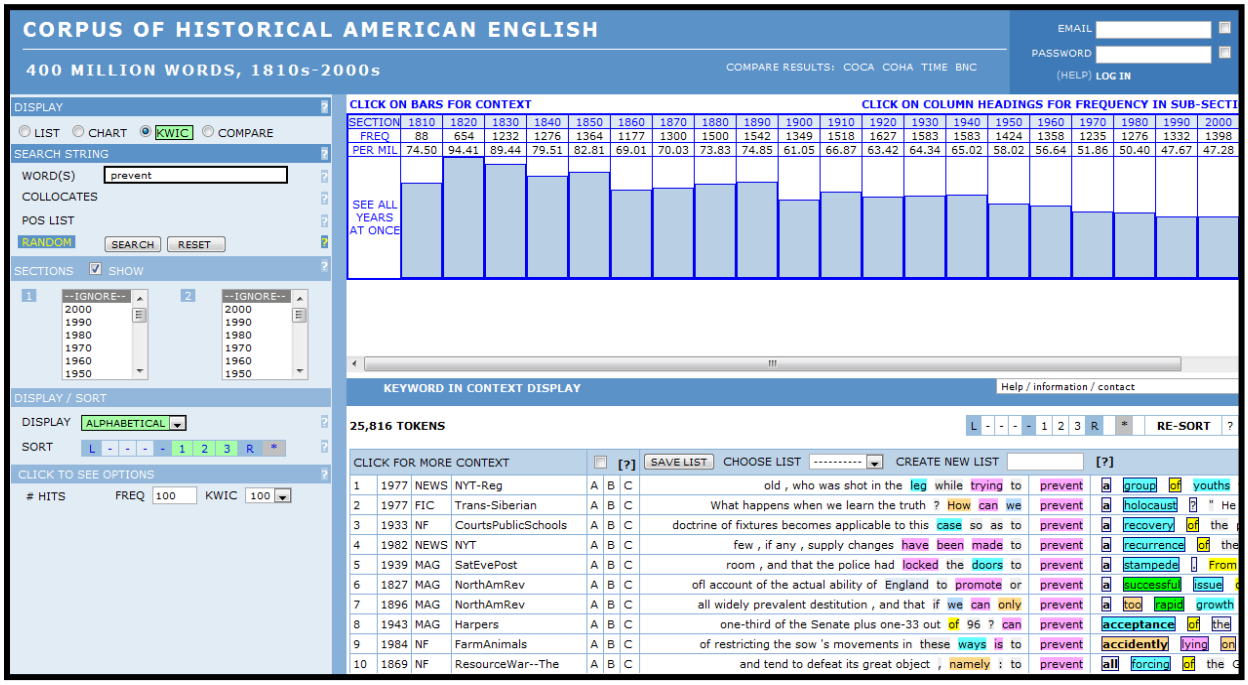
> 'a pattern consisting of any inflectional form of *prevent*, followed by a *single article* or *determiner* or *adjective* or *number* or *pronoun* or *adverb* or *alphabet* and a *noun*, then by *from* and any *-ing* forms'.

For the other search queries, the meanings are similar to (70) except for the different grammatical features as specified in Table 11 above. The same process is repeated for other sets of complementation clauses.

## 4.4 Methodology

Figure 2 below is a screenshot of the friendly and easy-to-use web interface of COHA. It shows the main parts of this interface. Based on it, the steps to collect the data are described.

Figure 2: Interface of COHA




As illustrated in Section 4.3, the search queries for all syntactic patterns were formed. There were a total of 19 different search queries because there were 19 different syntactic patterns for each set of complementation clauses (*prevent / stop + from* and *prevent / stop – from*) (refer Table 11 above in Section 4.3). A search query was entered in the box entitled 'WORD(S)' on the interface of COHA (left hand second row in Figure 2) to start the search. The 'CHART' button on the display (left hand top row in Figure 2) was selected to obtain the frequencies. A pop-up screen containing the frequencies in instances (raw figures) and in per million words appeared (see 'CLICK ON BARS FOR CONTEXT' in Figure 2). A total of 380 figures of raw

frequencies (20 decades with 19 search queries for each decade) were collected by the end of the process. After listing down all the raw frequencies for each decade, they were summed up and converted to per million words for standardisation. These results were exported to Microsoft Excel. Further analysis and visualisation of results was carried out in statistical programming environment R (refer http://www.r-project.org/).

To view and extract the concordances of the *prevent* and *stop* complementation clauses, the search query was again entered in the box entitled 'WORD(S)'. The 'KWIC' button was selected instead (left hand top row in Figure 2). A pop-up screen containing the concordances with highlighted keywords appeared (see 'KEYWORD IN CONTEXT DISPLAY' in Figure 2). The list of concordances for that particular search query was extracted into a spreadsheet. The process was repeated with 19 different search queries for each set of complementation clauses. All the lists of concordances were further manipulated for analysis.

Certain lexis in the list of concordances needed cautious categorisation such as the *-ing* forms and noun phrases. After extraction, I copied the concordances into EditPad to separate the *-ing* forms and noun phrases. EditPad[5] has simplified this process as I was able to use regular expressions for the separation process. The *-ing* forms were detached from the list and were kept in a new column. They were used for the diachronic distinctive collexeme analysis based on Hilpert's methodology (2006). This methodology for the second part of the study will be discussed separately in Chapter 6.

[5] EditPad is a compact general-purpose text editor (refer http://www.editpadlite.com/).

For the noun phrases categorisation, the noun phrases from the list were first separated in a new column after extraction. To calculate the length of words within noun phrases, the number of words within the noun phrases in each decade was summed up. Then, this figure was divided by the number of noun phrases within that particular decade. With those figures obtained from the calculations, I plotted a scatter graph. These steps were repeated for each set of complementation clauses. The noun phrases were used to test Rohdenburg's Complexity Principle (1996) in Chapter 5.

Those concordances which have meanings that were not related to the study were eliminated. Hence, the results were not contaminated. Even if there are some which remains, they are unlikely to have any effect on the results as the sample is big (refer Section 5.2.1, Chapter 5). All the steps mentioned above have been carried out carefully and the data have been checked to avoid errors. This process, however, was associated with some limitations. These limitations are discussed in Chapter 7.

## 4.5 Summary

This chapter listed and explained the search queries which were formed for the extraction of data. It then discussed the methodology that was used in this study. It detailed each step which was taken for the process of the extraction of complementation clauses from COHA. In the following chapter, I will discuss the results of three different analyses of the complementation clauses – (1) overall frequency per decade, (2) proportional values, and (3) length of noun phrases.

# CHAPTER 5

# RESULTS OF ANALYSES

## 5.1 Overview

This chapter presents a quantitative analysis of the data obtained from the Corpus of Historical American English (COHA) across a period of 200 years in American English.

The first section (5.2) examines the distribution of frequencies of the *prevent* and *stop* complementation clauses which is divided into two parts. Section 5.2.1 reveals the results of the overall frequency per decade while Section 5.2.2 compares the proportional values between the with *from* variant and the without *from* variant. These results provide an indication of the development for both *prevent* and *stop* complementation clauses. Next is an analysis of the length of noun phrases based on the average number of words found within the noun phrases (Section 5.3). This analysis was conducted based on Rohdenburg's Complexity Principle (1996) as mentioned in Section 2.4 (Chapter 2). Subsequent section (5.3.1) is a discussion on a pattern of linguistic change in progress, densification of content which is related to the length of noun phrases. For all the analyses mentioned (overall frequency per decade, proportional values and length of noun phrases), I have included my own interpretation of the results. The final section (5.4) presents a summary of the chapter.

## 5.2 Distribution of frequencies

This section attempts to examine the distribution of frequencies of the *prevent* and *stop* complementation clauses to shed light on the following research questions:

(i) What frequency changes can be observed in 19th, 20th and 21st century American English for:

- *prevent / stop* + noun phrase + *from* + *-ing* (e.g. She *prevented / stopped it from eating.*)
- *prevent / stop* + noun phrase + *-ing* (e.g. She *prevented / stopped it eating.*)

(ii) Do changes in overall frequency reflect changes in preferences for the with *from* variant and the without *from* variant?

My prediction in Chapter 2 for the *prevent* and *stop* complementation clauses are both of them will demonstrate higher frequencies with *from* than without *from* as American English tends to favour only one variant based on previous studies. *Prevent NP from -ing* will observe a more stabilised trend as its development started earlier than *stop NP from -ing*. *Stop NP from -ing* will increase in frequency slowly as the trend develops further. The reason may be due to the differences in meanings and senses as *stop* has a wider range of meanings as stated in Section 2.3.1 (Chapter 2).

Based on my prediction, two analyses of the distribution of frequencies were conducted. The following sections discuss the results of the analyses. They are divided into two parts:

(i) overall frequency per decade for the *prevent* and *stop* complementation clauses (with and without *from* variants) – Section 5.2.1

(ii) comparison of the proportional values between the with *from* variant and the without *from* variant for *prevent* and *stop* – Section 5.2.2

### 5.2.1 Distribution of the overall frequency per decade

Tables 12 and 13 below present the overall frequency per decade for both the *prevent* and *stop* complementation clauses in standardised form (per million words – pmw). From the observation of the two tables, the frequencies for both *prevent* complementation clauses (with and without *from* variants) are larger than both *stop* complementation clauses because there are more occurrences of related *prevent* constructions than unrelated *stop* constructions in COHA (refer to the raw frequency – number of instances). *Stop* has wider meanings and senses; thus, certain *stop* constructions were not extracted for this analysis. This includes embedded clauses within the complementation clauses which were not caught by the search queries that were formed. These limitations will be discussed in Chapter 7 as methodological issues.

Graphical illustrations are used to provide clearer explanation on how the *prevent* and *stop* complementation clauses have developed over time. Based on Table 12 and 13, four scatter graphs were plotted as shown in Figures 3 and 6 below. Each graph represents an overall development of the *prevent* and *stop* complementation clauses individually (the with and without *from* variants). The *x*-axis represents the time periods from 1810 to 2009 (year) and the *y*-axis

represents the observed frequency of the constructions (per million words). Note that all the graphs were not plotted to similar scale due to the differences in frequencies.

Regression analysis was used to determine the strength of the relationship between the dependent variable (observed frequency of the constructions – per million words) and the series of the changing variable (time periods – year). The purpose of using regression analysis is to obtain a general model of the data structure and to make predictions for future observation (Howitt and Cramer, 2011: 87). A regression line (the best-fit line) was drawn in each scatter plot. The regression line is explained through the slope of the line ($\beta$) and the point where the line cuts at the vertical axis (intercept), along with *t*-test and the corresponding significance level. The types of regression analysis which were used for the graphs below are linear regression (straight line) and non-linear regression (exponential and polynomial).

Below is Table 12 and Figure 3 (Graphs 1 and 2) which illustrate the overall frequency distribution of the *prevent* and *stop NP from -ing* variants. Figure 4 and 5 illustrate the summary output of statistical calculation from R for Graphs 1 and 2.

Table 12: Distribution of *prevent NP from -ing* and *stop NP from -ing* (1810-2009): raw frequencies in instances and frequencies per million words

| Year | Corpus size | *prevent NP from -ing* | | *stop NP from -ing* | |
|---|---|---|---|---|---|
| | | Raw frequency (instances) | Frequency (pmw) | Raw frequency (instances) | Frequency (pmw) |
| 1810 | 1181205 | 18 | 15.24 | 0 | 0 |
| 1820 | 6927173 | 246 | 35.51 | 2 | 0.29 |
| 1830 | 13774588 | 407 | 29.55 | 5 | 0.36 |
| 1840 | 16048393 | 367 | 22.87 | 4 | 0.25 |
| 1850 | 16471649 | 399 | 24.22 | 5 | 0.30 |
| 1860 | 17054872 | 394 | 23.10 | 4 | 0.23 |
| 1870 | 18562265 | 452 | 24.35 | 12 | 0.65 |
| 1880 | 20315965 | 479 | 23.58 | 10 | 0.49 |
| 1890 | 20600843 | 472 | 22.91 | 15 | 0.73 |
| 1900 | 22097593 | 433 | 19.50 | 25 | 1.13 |
| 1910 | 22700638 | 462 | 20.35 | 31 | 1.37 |
| 1920 | 25653746 | 547 | 21.32 | 58 | 2.26 |
| 1930 | 24602615 | 474 | 19.27 | 60 | 2.44 |
| 1940 | 24347955 | 494 | 20.29 | 69 | 2.83 |
| 1950 | 24544831 | 450 | 18.33 | 83 | 3.38 |
| 1960 | 23977232 | 446 | 18.60 | 131 | 5.46 |
| 1970 | 23815191 | 470 | 19.74 | 143 | 6.00 |
| 1980 | 25315978 | 506 | 19.99 | 171 | 6.75 |
| 1990 | 27941535 | 572 | 20.47 | 278 | 9.95 |
| 2000 | 29567390 | 609 | 20.60 | 354 | 11.97 |

Figure 3: Overall diachronic development of *prevent NP from -ing* and *stop NP from -ing* (1810-2009): frequencies per million words

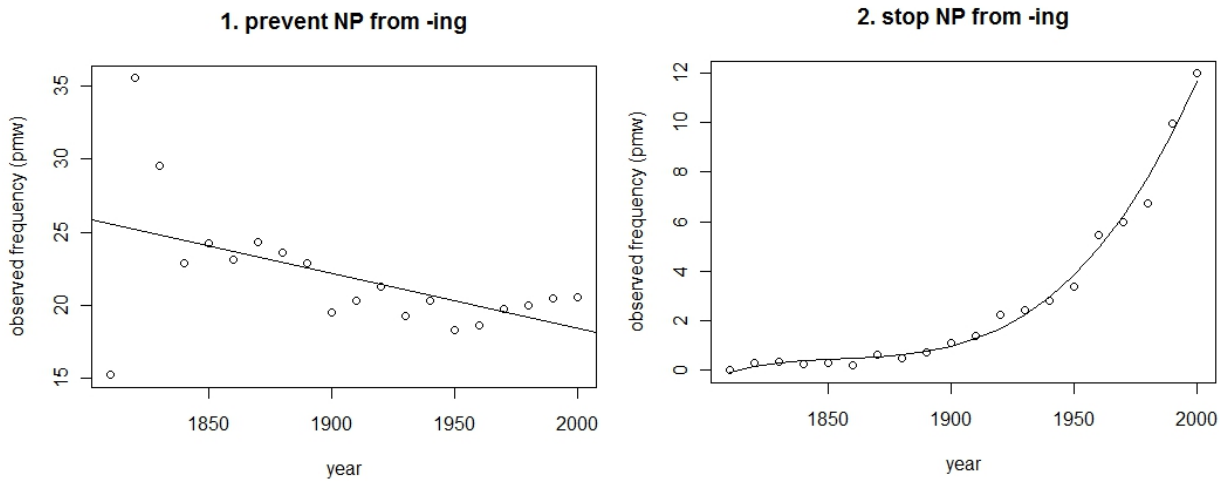


Figure 4: Summary output for *prevent NP from -ing* (Graph 1)

$Y = 93.115 – 0.037X$

```
lm(formula = pattern$prevent ~ pattern$year)

Residuals:
    Min      1Q  Median      3Q     Max
-10.2964  -1.4762  0.0338  0.8640  10.3469

Coefficients:
             Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  93.11475  28.44042    3.274    0.00422 **
pattern$year -0.03734   0.01492    -2.502   0.02222 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.848 on 18 degrees of freedom
Multiple R-squared: 0.258, Adjusted R-squared: 0.2168
F-statistic:  6.26 on 1 and 18 DF, p-value: 0.02222
```

Figure 5: Summary output for *stop NP from -ing* (Graph 2)

$Y = 2.842 + 13.130X + 6.867X^2 + 2.329X^3$

```
lm(formula = pattern$stop ~ poly(pattern$year, degree = 3))

Residuals:
    Min      1Q  Median      3Q     Max
-1.01336 -0.14918  0.05791  0.15962  0.56271

Coefficients:
                                 Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)                       2.84200   0.08656    32.834   4.14e-16 ***
poly(pattern$year, degree = 3)1  13.13034   0.38709    33.921   2.47e-16 ***
poly(pattern$year, degree = 3)2   6.86677   0.38709    17.739   6.03e-12 ***
poly(pattern$year, degree = 3)3   2.32922   0.38709     6.017   1.79e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3871 on 16 degrees of freedom
Multiple R-squared: 0.9895, Adjusted R-squared: 0.9875
F-statistic: 500.5 on 3 and 16 DF, p-value: 5.075e-16
```

Based on Table 12 above, for *prevent NP from -ing*, the overall frequencies are at a constant average of 21 per million words although they fluctuate from the start of the 19th century to early 21st century. The highest frequency lies in 1820 – 35 per million words. For *stop NP from -ing*, the trend is completely different from *prevent NP from -ing*. It starts off with very low frequencies (0 per million words) in the 19th century but gradually increases in the 20th century (2 per million words) and tops in the beginning of the 21st century (11 per million words).

Figure 3 above indicates that the overall diachronic trend for *prevent NP from -ing* is decreasing towards the 20th century but is increasing back slowly in the beginning of the 21st century (refer Graph 1) meanwhile for *stop NP from -ing*, it is increasing steadily towards the 21st century (refer Graph 2). In Graph 1, the observed frequency of *prevent NP from -ing* (per million words) significantly predicts the time periods (year), $\beta$ = -0.037, $t(18)$ = -2.502, $p$ <.05. The observed frequency of *prevent NP from -ing* (per million words) also explains a small proportion of variance in time periods (year), adjusted $R^2$ = 0.219, $F(1,18)$ = 6.26, $p$ <.05. In Graph 2, the observed frequency of *stop NP from -ing* (per million words) significantly predicts the time periods (year), $\beta_1$ = 13.130, $t(16)$ = 33.921, $p$ <.001, $\beta_2$ = 6.867, $t(16)$ = 17.729, $p$ <.001 and $\beta_3$ = 2.329, $t(16)$ = 6.017, $p$ <.001. The observed frequency of *stop NP from -ing* (per million words) also explains a significant proportion of variance in time periods (year), adjusted $R^2$ = 0.988, $F(3,16)$ = 500.5, $p$ <.001.

The overall frequencies for *prevent NP from -ing* are high despite the variant has a low adjusted $R^2$. This phenomenon may gives an indication that the with *from* variant is still being favoured

very much in present-day American English. This result is similar to the result that was obtained by Mair (2002: 112) which show high instances of *prevent NP from -ing* in American English. On the other hand, the overall frequencies for *stop NP from -ing* increase exponentially over time. The adjusted $R^2$ value is also very high. This means that *stop* has a strong relationship with the preposition or complementiser *from* in complementation clauses. *Stop NP from -ing* may continue to double its frequencies in the future. Hence, it needs further investigation.

As discussed in Section 2.2.2 (Chapter 2), Huddleston and Pullum (2002: 657) describe the function of *from* as a role of 'intended actions [which act] as spatial goals' and explain that 'to hold someone back from doing something means to hold them back in order for them not to perform the action'. Rohdenburg (2009: 211) adds that *from* plays a special role in verbs of negative causation such as *prevent* and *stop* in complex American English constructions. These claims are supported by the examples which I have found in COHA.

(71) Stone walls, electric, fences, cameras, motion, detectors, security guards, all for the purpose of *preventing an enterprising lawbreaker <u>from</u> climbing* over the walls into the sanctity of private space beyond. [Enoch, 2005, *Flirting with Danger* (Fiction)]

(72) As a show of faith, it's a tough act to follow but that has *stopped some ingenious American hucksters <u>from</u> trying*. [Rich, 2005, *A High-Tech Lynching in Prime Time* (News)]

In (71), the *lawbreaker* which is denoted by the object noun phrase of *preventing*, may be able to keep a distance from the situation of *climbing over the walls*, which is a prevented event denoted by the *-ing* clause. This phenomenon is similar with (72) in which *trying* implies less power on the subject of the *-ing* clause (*hucksters*). Both noun phrases are holding back their actions so that these actions will not be performed. This notion may be related to verbs implying non-agentive on their subject noun phrases and have less responsibility as an agent role.

In Section 2.3.2 (Chapter 2), Sellgren (2010: 49) discusses the idea of hypotheticality which is indicated by the *-ing* forms. She mentions that a difference in syntactic form entails a difference in meaning. This idea is applicable to examples like (73) and (74) which are obtained from COHA.

(73) He almost felt as if by taking the goose away he was *preventing his parents <u>from</u> committing* a crime, for it could hardly be less than one to kill so intelligent and loving a creature. [Various, 1886, *The Little Gold Miners of the Sierras and Other Stories* (Fiction)]

(74) In 1875, Congress passed a law *stopping Chinese prostitutes <u>from</u> coming* to the United States, while the same class of Caucasian origin were and are now brought to San Francisco. [Williams, 1879, *Our Treaties with China* (Magazine)]

In (73), the preventer feels that if he takes the goose away, he is trying to *prevent his parents from committing a crime*. The difference in meaning for (73) is indicated by *if*. In (74), the United States wanted to *stop the Chinese prostitutes from entering* the country and thus, a law

was passed but the event has actually taken place in the past (before 1875). It was indicated by the usage of past tense (*were brought*) and present tense (*are now brought*). Both examples appear to follow the semantic distinction which derives from the idea of hypotheticality.

Below is Table 14 and Figure 6 (Graphs 3 and 4) which illustrate the overall frequency distribution of the *prevent* and *stop NP -ing* variants. Figure 7 and 8 illustrate the summary output of statistical calculation from R for Graphs 3 and 4.

Table 13: Distribution of *prevent NP -ing* and *stop NP -ing* (1810-2009): raw frequencies in instances and frequencies per million words

| Year | Corpus size | *prevent NP -ing* | | *stop NP -ing* | |
|---|---|---|---|---|---|
| | | Raw frequency (instances) | Frequency (pmw) | Raw frequency (instances) | Frequency (pmw) |
| 1810 | 1181205 | 1 | 0.85 | 0 | 0 |
| 1820 | 6927173 | 27 | 3.90 | 1 | 0.14 |
| 1830 | 13774588 | 47 | 3.41 | 1 | 0.07 |
| 1840 | 16048393 | 54 | 3.36 | 2 | 0.12 |
| 1850 | 16471649 | 70 | 4.25 | 4 | 0.24 |
| 1860 | 17054872 | 75 | 4.40 | 4 | 0.23 |
| 1870 | 18562265 | 93 | 5.01 | 6 | 0.32 |
| 1880 | 20315965 | 81 | 3.99 | 12 | 0.59 |
| 1890 | 20600843 | 84 | 4.08 | 15 | 0.73 |
| 1900 | 22097593 | 107 | 4.84 | 12 | 0.54 |
| 1910 | 22700638 | 89 | 3.92 | 26 | 1.18 |
| 1920 | 25653746 | 88 | 3.43 | 24 | 0.97 |
| 1930 | 24602615 | 55 | 2.28 | 30 | 1.26 |
| 1940 | 24347955 | 38 | 1.60 | 21 | 1.27 |
| 1950 | 24544831 | 40 | 1.63 | 29 | 1.22 |
| 1960 | 23977232 | 42 | 1.75 | 37 | 1.63 |
| 1970 | 23815191 | 42 | 1.76 | 49 | 2.10 |
| 1980 | 25315978 | 22 | 1.11 | 37 | 1.54 |
| 1990 | 27941535 | 22 | 0.79 | 33 | 1.18 |
| 2000 | 29567390 | 18 | 0.64 | 35 | 1.18 |

Figure 6: Overall diachronic development of *prevent NP -ing* and *stop NP -ing* (1810-2009): frequencies per million words

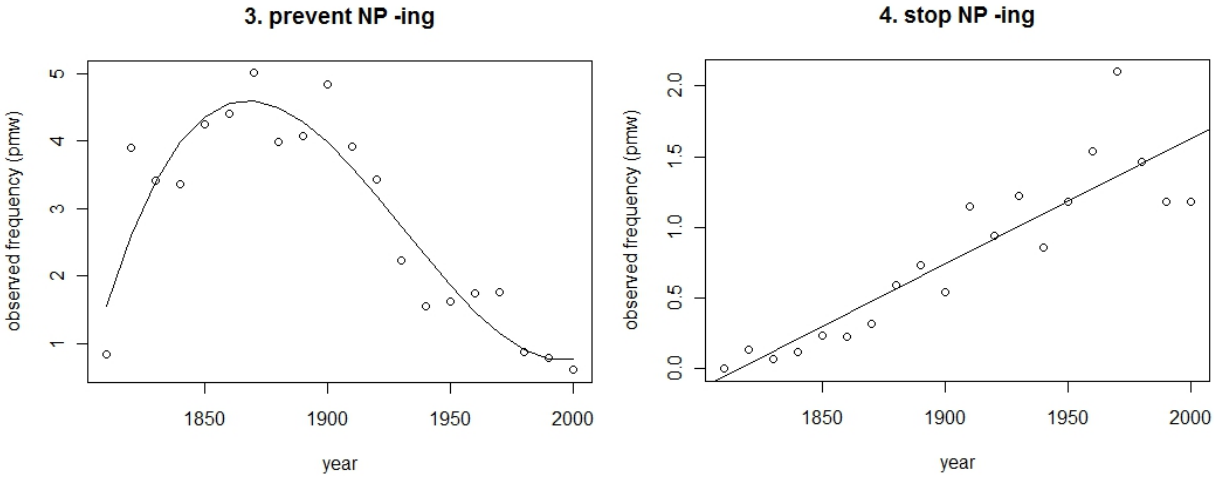


Figure 7: Summary output for *prevent NP -ing* (Graph 3)

$Y = 2.833 – 3.966X – 3.878X^2 + 2.450X^3$

```
lm(formula = pattern$prevent ~ poly(pattern$year, degree = 3))

Residuals:
    Min      1Q      Median      3Q      Max
-0.73700  -0.30425  -0.07721  0.28094  1.29677

Coefficients:
                                  Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)                        2.8325    0.1286     22.031   2.14e-13 ***
poly(pattern$year, degree = 3)1   -3.9657    0.5750     -6.897   3.59e-06 ***
poly(pattern$year, degree = 3)2   -3.8778    0.5750     -6.744   4.71e-06 ***
poly(pattern$year, degree = 3)3    2.4500    0.5750      4.261   0.000597 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.575 on 16 degrees of freedom
Multiple R-squared: 0.8742, Adjusted R-squared: 0.8506
F-statistic: 37.07 on 3 and 16 DF, p-value: 1.970e-07
```

Figure 8: Summary output for *stop NP -ing* (Graph 4)

$Y = -16.070 + 0.009X$

```
lm(formula = pattern$stop ~ pattern$year)

Residuals:
    Min      1Q      Median      3Q      Max
-0.45029  -0.16011 -0.00052   0.08064  0.73523

Coefficients:
                 Estimate     Std. Error  t value  Pr(>|t|)
(Intercept)    -16.070466   1.966068   -8.174   1.80e-07 ***
pattern$year   0.008850     0.001032    8.579   8.93e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.266 on 18 degrees of freedom
Multiple R-squared: 0.8035, Adjusted R-squared: 0.7926
F-statistic: 73.61 on 1 and 18 DF, p-value: 8.928e-08
```

Based on Table 13 above, for *prevent NP -ing*, the overall frequencies are much lesser than those of *prevent NP from -ing*. They are only one sixth of the overall frequencies of the with *from* variant (3 per million words in comparison to 20 per million words). The numbers of constructions are low based on the raw frequencies of *prevent NP -ing*. For *stop NP -ing*, the trend remains the same where it begins with low frequencies in the 19th century (0 per million words) and these values gradually increase towards the 21st century (1 per million words). However, the only difference is at the beginning of the 21st century when the frequencies are lesser in comparison to the with *from* variant (1 per million words in comparison to 11 per million words).

Figure 6 above indicates that the overall diachronic trend for *prevent NP -ing* is decreasing tremendously especially at the start of the 20th century and continues to decrease towards the 21st century (refer Graph 3). Meanwhile, *stop NP -ing* is increasing steadily at first but starts to decrease from 1960s (refer Graph 4). In Graph 3, the observed frequency of *prevent NP -ing* (per million words) significantly predicts the time periods (year), $\beta_1$ = -3.966, $t(16)$ = -6.987, $p$ <.001, $\beta_2$ = -3.878, $t(16)$ = -6.744, $p$ <.001 and $\beta_3$ = 2.450, $t(16)$ = 4.261, $p$ <.001. The observed frequency of *prevent NP -ing* (per million words) also explains a significant proportion of variance in time periods (year), adjusted $R^2$ = 0.851, $F(3,16)$ = 37.07, $p$ <.001. In Graph 4, the observed frequency of *stop NP -ing* (per million words) significantly predicts the time periods (year), $\beta$ = 0.009, $t(18)$ = 8.579, $p$ <.001. The observed frequency of *stop NP -ing* (per million words) also explains a significant proportion of variance in time periods (year), adjusted $R^2$ = 0.794, $F(1,18)$ = 73.61, $p$ <.001.

The overall frequencies for *prevent NP -ing* are low but the variant has a high adjusted $R^2$. This result of low overall frequencies is similar to the result by Mair (2002: 112) which shows low instances for *prevent NP -ing*. As discussed above, the with *from* variant is increasing and being favoured at present day. As a result, the without *from* variant will continue to decrease and will be used less in American English. Although the adjusted $R^2$ is high, the overall frequencies for *stop NP -ing* are lower than *prevent NP -ing*. This may be due to the unstable development of the *stop* complementation clauses which causes *stop NP -ing* to have low frequencies. This variant needs more attention for further investigation.

In Section 2.3.2 (Chapter 2), Rudanko (2002: 58) and Dixon (1995: 217) comment that the without *from* variant implies a sense of directness, immediacy and external observability between the preventer and preventee. Although this variant is rare in American English, it can still be found in COHA. Below are some examples from COHA:

(75) When output is falling and unemployment rising, these payments *prevent disposable incomes falling* as fast as factor incomes. [Stanlake, 1979, Macro-economics: An Introduction (Academic)]

(76) The Behaviorists say they are bad, not mad, and we can *stop them being* bad by utilizing new techniques. [Mitford, 1973, *The Torture Cure* (Magazine)]

In (75), the action of *preventing incomes falling* which is caused by the subject noun phrase of *prevent*, *payments* employs direct means of prevention. In (76), the action of *stopping* by the preventer, *we* denotes a direct sense of immediacy and directness to the preventee, *behaviourists*

from *being bad*. This notion of directness is only applicable to animate and non-abstract object noun phrases of the complementation clauses like *incomes* and *them* (*the behaviourists*).

According to Quirk et al. (1985: 1194) and Aarts (1990: 152) in Section 2.2.2 (Chapter 2), the element *from* acts as a complementiser in complementation clauses. A complementiser is meaningless in many constructions and does not carry any importance as this is a characteristic of most complementisers. For this reason, *from* is optional especially in active complementation clauses. This applies to examples (77) and (78) which are found in COHA.

(77) We know not our case is affected by this inconsistency, or that is does *prevent a suit being* brought before the court to liberate a man held as a slave; and therefore we shall leave it for our lawyers to speculate upon. [Mellen, 1841, *An Argument on the Unconstitutionality of Slavery, Embracing an Abstract of the National and State Conventions on this Subject* (Academic)]

(78) Subsidiary to this principal undertaking, a force of heavy armed boats was to be maintained on the St. Lawrence, to *stop the British supplies arriving* from Montreal; and a minor expedition, naval and military, was to be sent against Mackinac. [Mahan, 1905, *The War of 1812* (Magazine)]

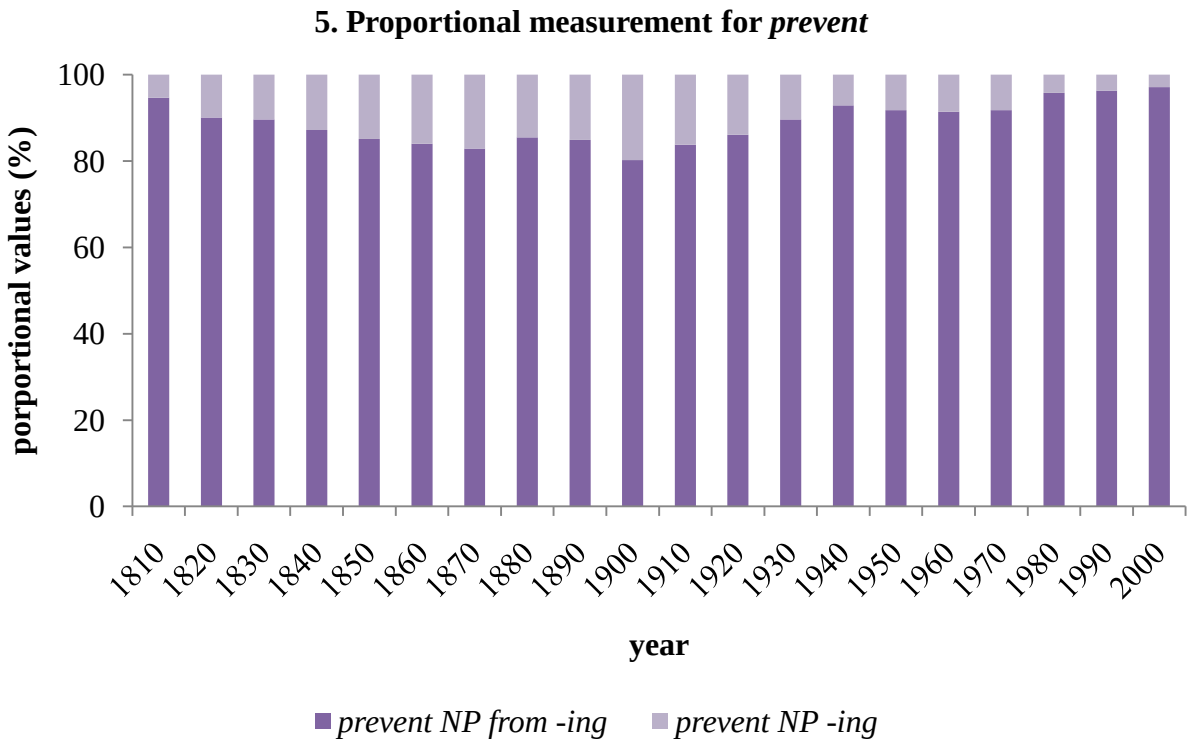
In (77) and (78), the actions of *preventing a suit being bought* and *stopping the British supplies arriving* can omit the complementiser, *from* in their active complementation clauses. The meaning of the actions remains the same even without the complementiser, *from*. The passive complementation clauses will not be discussed as the search queries formed did not include them.

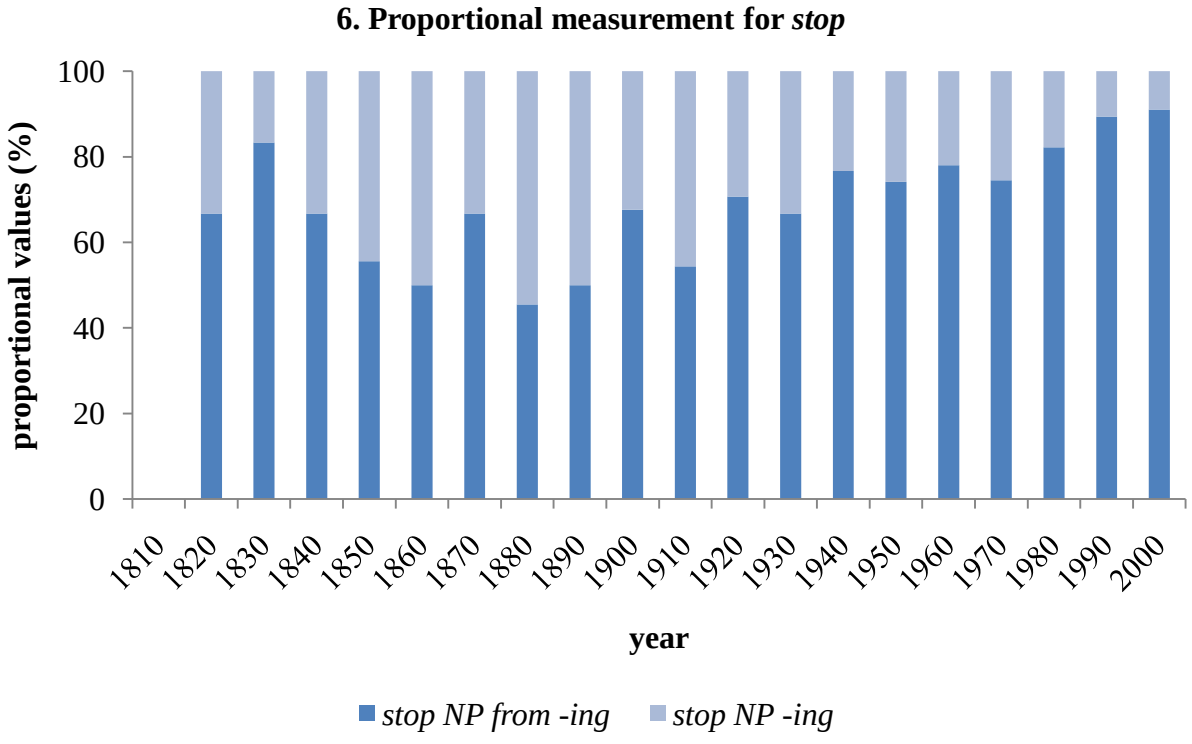
In summary, based on the overall quantitative results above, both *prevent* and *stop* complementation clauses appear opposite when compared against one another. In general, both the *from* variants for *prevent* and *stop* are increasing while both the without *from* variants are decreasing. *Prevent NP from -ing* is increasing slower than *stop NP from -ing* based on Figure 3. In contrast, *prevent NP -ing* is decreasing faster than *stop NP -ing* based on Figure 6. This may give a hint that the *prevent* complementation clauses are more developed in comparison to the *stop* complementation clauses.

### 5.2.2 Distribution of the proportional values

As the analysis results in Section 5.2.1 reveal that the overall diachronic trends have demonstrated unstable movements, it is important to find out which variant (with or without *from*) is being favoured more in 19$^{th}$, 20$^{th}$ and 21$^{st}$ century American English. Hence, a proportional measurement is used for this observation. The raw frequencies for *prevent NP from -ing* and *prevent NP -ing* were calculated in proportions. These proportions were then converted to percentage for standardisation purpose as the corpus size for each decade is different. With these percentages, a single graph was plotted for both *prevent* complementation clauses as shown in Figure 9 (Graph 5) below. The *x*-axis represents the time periods from 1810 to 2009 (year) and the *y*-axis represents the proportional values (%). These steps were also repeated for both *stop* complementation clauses and the result is demonstrated in Figure 9 (Graph 6) below. Both graphs were plotted to similar scale as the maximum proportional value is 100%.

Figure 9: Proportional measurement for the *prevent* and *stop* complementation clauses: frequencies in percentage (%)



Based on Figure 9 (Graph 5) above, it is undeniable that *prevent NP from -ing* is dominating from 19th to 21st century American English when compared to *prevent NP -ing*. The without *from* variant (light purple coloured bars) is almost an average of a tenth in proportion of the with *from* variant (dark purple coloured bars) starting from the 19th century until the beginning of the 21st century. Although the proportional values for the with *from* variant decrease towards 1900, they are still approximately an eighth of the total proportion when compared to the opposite variant. Since then, these values start to increase again and stabilise towards 2000. These high ratio values of favouring the with *from* variant typically show a tendency of omitting the usage of the without *from* variant completely in the late 21st century (as shown by the dark purple coloured bars). It is predicted that *prevent NP from -ing* may take over *prevent NP -ing* in American English and become a dominant trend in the future. However, this prediction could possibly not be accurate and will need more investigation because language never stops developing.

As Figure 9 (Graph 6) above illustrates, *stop NP from -ing* is being favoured when compared to *stop NP from -ing* in American English. The proportional values for the without *from* variant (light blue coloured bars) is almost half of the proportional values for the with *from* variant (dark blue coloured bars) at the beginning of the 19th century although they fluctuate. In the early 20th century, the proportional values of the with *from* variant start to increase. The proportional values for the with *from* variant are an average of an eighth of the total proportion towards the beginning of the 21st century. This indicates that *stop NP from -ing* has a strong tendency to become a more preferred variant than *stop NP -ing* in American English at the beginning of the 21st century due to the increasing proportional values as demonstrated in Figure 9 (Graph 6) above. It is also predicted that this trend will continue to dominate in the 21st century like *prevent*

*NP from -ing*. However, the prediction might not be precise as the development of these complementation clauses may change in the future.

Hilpert and Gries (2009: 386) comment that trends in diachronic historical corpora are not unidirectional or not strong enough to provide clear-cut indication to the process of language change. Trends may increase or decrease in strength over time. The basic observation should focus on whether a trend has become more or less common, or whether it remains stabilised. However, it is never obvious to observe a trend which constitutes a significant development or an accidental fluctuation in the data.

Both graphs above do not display stabilised trends but trends which are more and less common at different time periods. Although the with *from* variant for *prevent* dominates from 1810 to 2009, it decreases and reaches its lowest frequency in 1900. It then gradually increases and becomes a common trend towards the 21st century. The with *from* variant for *stop* fluctuates from 1810 to 1920. It does not demonstrate a clear-cut indication on how the trend will develop. However, after 1920, it increases slowly and becomes more common towards the 21st century. Both phenomena give an indication that American English keeps on developing because they do not show stable directions. Further investigation is still needed for deeper exploration on both complementation clauses.

In short, from Figure 9 above, we observe that the *prevent* and *stop NP from -ing* variants are being used more often than the *prevent* and *stop NP -ing* variants in American English for all three centuries. The development for *prevent NP from -ing* has almost come to a halt when compared to *stop NP from -ing* as its overall trend is more stabilised. Its proportional values are also much higher and more consistent in comparison with the proportional values for *stop NP from -ing*. Towards the 21st century, there is very little usage of *prevent NP -ing*. On the other hand, the trend for s*top NP from -ing* is still increasing and is expected to behave like *prevent NP from -ing* which is to neglect the usage of *stop NP -ing* in the future.

## 5.3 Distribution of the length of noun phrases

After the analyses of the overall development of the complementation clauses, another analysis of the length of noun phrases was conducted to test Rohdenburg's Complexity Principle (1996). The research question which was formed at the beginning of the study is as follows:

(i) Is there a connection between the length of noun phrases and the use of preposition or complementiser *from*?

My hypothesis is the length of noun phrases increases when the degree of grammatical explicitness in the *prevent* and *stop* complementation clauses increases. The preference for *from* to appear in the complementation clauses should coincide with longer noun phrases. This is reflected in the average number of words within the noun phrases which were extracted from the samples. The average number of words for the with *from* variant should be higher than the without *from* variant for both *prevent* and *stop* because they have longer and more complex noun

descriptions. As time moves towards the 21st century, the length of noun phrases should increase too. More information is compressed in longer noun phrases to provide more efficient communication especially in newspaper language.

Figure 10 (Graphs 7, 8, 9 and 10) below illustrates the average number of words within the noun phrases for both *prevent* and *stop* complementation clauses. These graphs are used for the comparison of the length of noun phrases between the with *from* variant and the without *from* variant. The *x*-axis represents the time periods from 1810 to 2009 (year) and the *y*-axis represents the average number of words within the noun phrases (words). All the graphs were not plotted to similar scale due to the differences in average number of words. Regression lines were drawn in each graph to determine the strength of the relationship between the average number of words within the noun phrases (words) and the time periods (year). The regression line is explained through the slope of the line ($\beta$) and the point where the line cuts at the vertical axis (intercept), along with *t*-test and the corresponding significance level. The type of regression analysis which was used for the graphs below is non-linear regression (power and polynomial). Figures 11, 12, 13 and 14 illustrate the summary output of statistical calculation from R for Graphs 7, 8, 9 and 10.

Figure 10: Average number of words within the noun phrases for the *prevent* and *stop* complementation clauses: frequencies in words

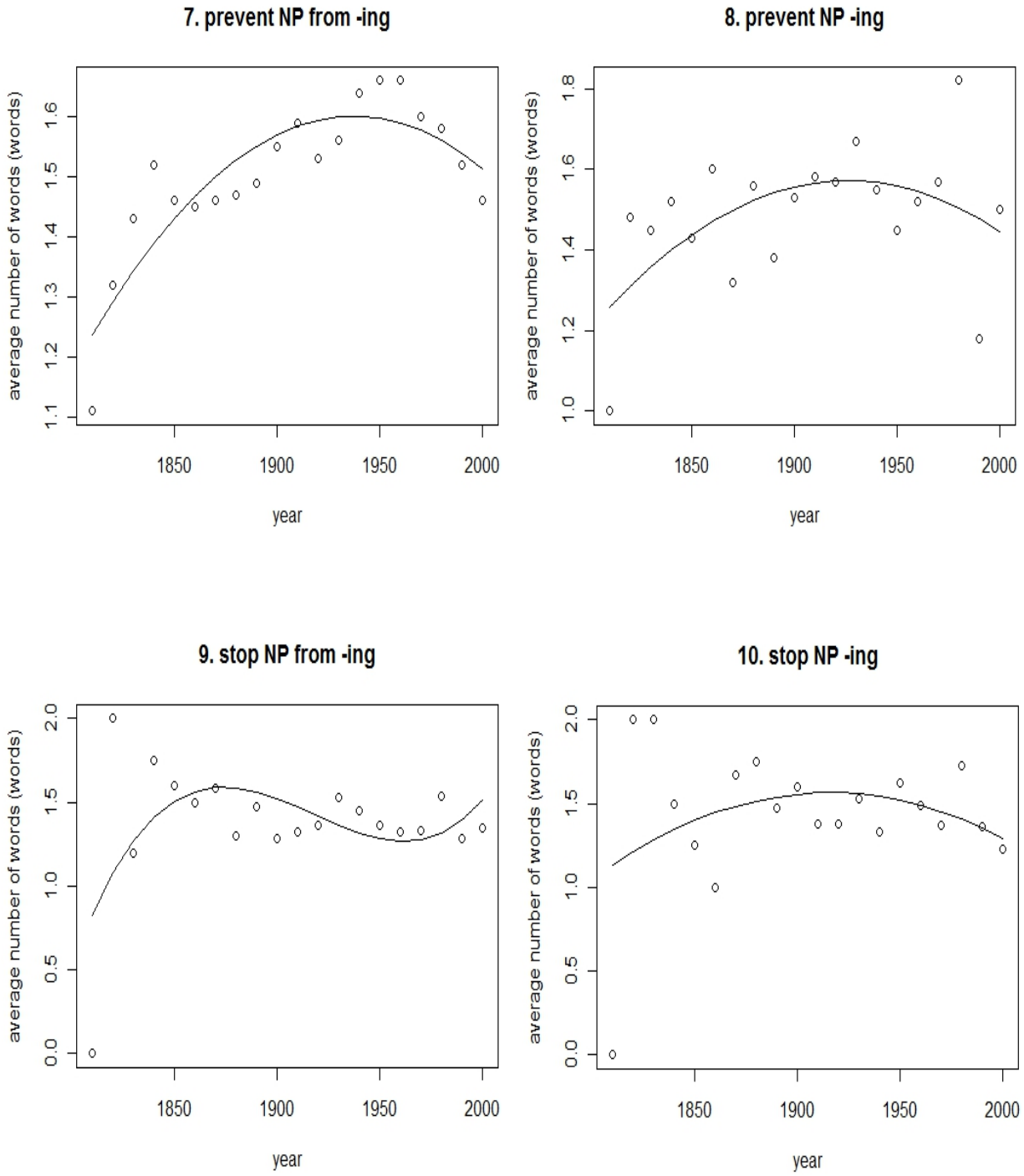


Figure 11: Summary output for *prevent NP from -ing* (Graph 7)

$Y = 1.503 + 0.374X - 0.297X^2$

```
lm(formula = pattern$prevent ~ poly(pattern$year, degree = 2))

Residuals:
     Min       1Q       Median       3Q      Max
-0.127286 -0.043132 -0.005569  0.032338 0.131511

Coefficients:
                                  Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)                        1.50300   0.01458   103.103  < 2e-16 ***
poly(pattern$year, degree = 2)1    0.37421   0.06519    5.740   2.40e-05 ***
poly(pattern$year, degree = 2)2 -0.29721   0.06519   -4.559   0.000278 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.06519 on 17 degrees of freedom
Multiple R-squared: 0.7597, Adjusted R-squared: 0.7314
F-statistic: 26.87 on 2 and 17 DF, p-value: 5.459e-06
```

Figure 12: Summary output for *prevent NP -ing* (Graph 8)

$Y = 1.484 + 0.252X - 0.308X^2$

```
lm(formula = pattern$prevent ~ poly(pattern$year, degree = 2))

Residuals:
     Min       1Q       Median       3Q      Max
-0.296575 -0.046414  0.006843  0.094349 0.316031

Coefficients:
                                  Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)                        1.48400   0.03477   42.676   <2e-16 ***
poly(pattern$year, degree = 2)1    0.25245   0.15551    1.623   0.1229
poly(pattern$year, degree = 2)2 -0.30793   0.15551   -1.980   0.0641 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1555 on 17 degrees of freedom
Multiple R-squared: 0.2783, Adjusted R-squared: 0.1934
F-statistic: 3.278 on 2 and 17 DF, p-value: 0.06251
```

Figure 13: Summary output for *stop NP from -ing* (Graph 9)

$Y = 1.376 + 0.168X – 0.482X^2 + 0.637X^3$

```
lm(formula = pattern$stop ~ poly(pattern$year, degree = 3))

Residuals:
    Min      1Q      Median      3Q     Max
-0.82775 -0.12184 -0.03241  0.10380 0.92406

Coefficients:
                                  Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)                       1.37600   0.07891     17.438   7.83e-12 ***
poly(pattern$year, degree = 3)1   0.16791   0.35289      0.476   0.6406
poly(pattern$year, degree = 3)2  -0.48227   0.35289     -1.367   0.1906
poly(pattern$year, degree = 3)3   0.63738   0.35289      1.806   0.0897 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3529 on 16 degrees of freedom
Multiple R-squared: 0.2508, Adjusted R-squared: 0.1103
F-statistic: 1.785 on 3 and 16 DF, p-value: 0.1904
```

Figure 14: Summary output for *stop NP -ing* (Graph 10)

$Y = 1.433 + 0.213X – 0.511X^2$

```
lm(formula = pattern$stop ~ poly(pattern$year, degree = 2))

Residuals:
    Min      1Q      Median      3Q     Max
-1.13443 -0.15794 -0.01398  0.16285 0.78783

Coefficients:
                                  Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)                       1.43300   0.09399     15.247   2.39e-11 ***
poly(pattern$year, degree = 2)1   0.21328   0.42032      0.507   0.618
poly(pattern$year, degree = 2)2  -0.51140   0.42032     -1.217   0.240
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4203 on 17 degrees of freedom
Multiple R-squared: 0.09274, Adjusted R-squared: -0.01399
F-statistic: 0.8689 on 2 and 17 DF, p-value: 0.4372
```

Based on Figure 10 (Graphs 7 and 8) above, it is observed that both *prevent* complementation clauses have almost similar average number of words within the noun phrases which is approximately 1.4 to 1.6 words. *Prevent NP from -ing* has a more consistent trend for the average number of words as the trend increases at the beginning of the 19th and 20th centuries although it decreases towards the 21st century (refer Graph 7). On the other hand, *prevent NP -ing* has a wild fluctuation trend (refer Graph 8). In Graph 7, the average number of words within the noun phrases for *prevent NP from -ing* (words) significantly predicts the time periods (year), $\beta_1 = 0.374$, $t(17) = 5.740$, $p < .001$ and $\beta_2 = -0.297$, $t(17) = -4.559$, $p < .001$. The average number of words within the noun phrases for *prevent NP from -ing* (words) also explains a significant proportion of variance in time periods (year), adjusted $R^2 = 0.731$, $F(2,17) = 26.87$, $p < .001$. In Graph 8, the average number of words within the noun phrases for *prevent NP -ing* (words) significantly predicts the time periods (year), $\beta_1 = 0.252$, $t(17) = 1.623$, $p < 1$ and $\beta_2 = -0.308$, $t(17) = -1.980$, $p < .1$. The average number of words within the noun phrases for *prevent NP -ing* (words) also explains a small proportion of variance in time periods (year), adjusted $R^2 = 0.193$, $F(2,17) = 3.278$, $p < .1$. Based on the observation of Figure 10 (Graphs 7 and 8), there is a tendency for the usage of longer noun phrases after the 20th century when compared both variants although this trend is not clearly seen. Longer noun phrases provide more comprehensive information and create more explicit constructions. Thus, the preposition or complementiser *from* is inserted into the complementation clauses after the long noun phrases.

From the observation of Graphs 9 and 10 in Figure 10 above, both *stop* complementation clauses have average number of words within the noun phrases which is between 1.4 to 1.8 words. *Stop NP from -ing* has a slightly more consistent trend than *stop NP -ing* as the average number of

words for *stop NP -ing* fluctuate more. The average number of words trend for *stop NP from -ing* is decreasing in general (refer Graph 9) as well as for *stop NP -ing* but the trend for *stop NP ing* is decreasing more (refer Graph 10). In Graph 9, the average number of words within the noun phrases for *stop NP from -ing* (words) significantly predicts the time periods (year), $\beta_1 = 0.168$, $t(16) = 0.476$, $p < 1$, $\beta_2 = -0.482$, $t(16) = -1.367$, $p < 1$ and $\beta_3 = 0.637$, $t(16) = 1.806$, $p < .1$. The average number of words within the noun phrases for *stop NP from -ing* (words) also explains a small proportion of variance in time periods (year), adjusted $R^2 = 0.110$, $F(3,16) = 1.785$, $p < 1$. In Graph 10, the average number of words within the noun phrases for *stop NP -ing* (words) significantly predicts the time periods (year), $\beta_1 = 0.213$, $t(17) = 0.507$, $p < 1$ and $\beta_2 = -0.511$, $t(17) = -1.217$, $p < 1$. The average number of words within the noun phrases for *stop NP -ing* (words) also explains a tiny proportion of variance in time periods (year), adjusted $R^2 = 0.093$, $F(2,17) = 0.869$, $p < 1$. From Figure 10 (Graphs 9 and 10), we can observe that *stop NP from -ing* is showing a tendency to increase back at the beginning of the 21st century while *stop NP -ing* continues to decrease. This trend actually supports Rohdenburg's Complexity Principle (1996). The more explicit options are observed to attract to the more complex noun phrases although they are not clearly shown with the *stop* complementation clauses. A possible reason to explain this phenomenon is that *stop* has fewer constructions based on the raw frequencies as shown in Tables 12 and 13 (Section 5.2.1) in comparison to *prevent* which might affect the results during the analysis.

According to the results from Figure 10 above, in general, both *prevent* and *stop* complementation clauses favour the preposition or complementiser *from* when they have long noun phrases in order to show explicitness in their constructions. Longer noun phrases are used

towards the 21st century too. As mentioned by Rohdenburg (1996: 151) in Section 2.4 (Chapter 2), the more explicit constructions will tend to be favoured in cognitively more complex environments. The with *from* variant is considered as complex environment for this study (cf. Sellgren, 2010: 49). A reason which is related in processing of the more or less explicit grammatical options as stated in Section 2.4 (Chapter 2) is working memory in the brain (cf. Hawkins, 1990 and 1992).

When processing linguistic expressions, there are various kinds of cognitive involvement between the speaker and the hearer. Working memory plays an important role in these complex linguistic processes (Aiello, 2010: 3). Long noun phrase increases simultaneous processing demands in working memory. This is because the more complex a subject is, the longer the brain takes to recognise its relationship and to identify the actual function of the linguistic expression. Working memory requires dependent processing across other properties in the linguistic expression which go beyond the processing of additional forms and meanings through explicit marking. As the importance of the long noun phrase increases during the processing of the relevant information, other properties within that linguistic expression will also increase simultaneously. However, it is the long noun phrase structure which needs more recognition for accomplishment. It is difficult to process the content of the long noun phrase instantly because it is packed with a sequence of nouns. Thus, there will be more demands in working memory to store that respective information from the linguistic expression and to process it. This means that more decisions have to be made before completing the processing of that linguistic expression.

Working memory is needed to remind the speaker and the hearer about the relevant information which is stored in the complementation clauses during the processing of the long noun phrases. The key information within the complementation clauses is the emphasis of the meaning of *prevent* or *stop* and not the other properties. This explains a possible reason for the insertion of the preposition or complementiser *from* after the long noun phrases within the complementation clauses. The usage of *from* also illustrates formality in those more explicit grammatical options than the less explicit counterpart. Long noun phrases are hardly found in the without *from* variant which is considered as simple environment. Hence, this phenomenon supports Rohdenburg's Complexity Principle.

In summary, the results from Figure 10 above demonstrate that the preposition or complementiser *from* plays an important role in influencing the explicitness of the constructions. The with *from* variant for both *prevent* and *stop* are determined by the complexity of the noun phrases preceding the gerund. The more explicit the constructions are with references to longer noun phrases, the higher the tendency for *from* is used. Nowadays, we can observe that longer noun phrases are used to provide much thorough information regarding the content of those noun phrases. *Prevent NP from -ing* support the hypothesis strongly as the average number of words is increasing in general when compared with *prevent NP -ing*. However, *stop NP from -ing* does not show strong evidence to support the hypothesis even though the average number of words starts to increase at the beginning of the 21$^{st}$ century. It needs further investigation.

### 5.3.1 Densification of content

The results from my analysis of the length of noun phrases in Section 5.3 show evidence which agree with the results from the study by Biber (2003) and Svartvik and Leech (2006). As time moves towards the 21st century, the length of noun phrases increases. This means that more nouns are compressed in a single noun phrase construction to create a long noun phrase which will provide more compact information. There is evidence from my analysis which show that long noun phrases are being used starting in the late 20th century within *prevent NP from -ing* and *stop NP from -ing*. These long noun phrases are completely absent in both *prevent NP -ing* and *stop NP -ing*. In *prevent NP -ing* and *stop NP -ing*, the noun phrases tend to be shorter. This allows the preventer to employ a sense of directness with the prevented actions.

Below are some examples of the long noun phrases found in my data extraction from COHA:

(79) If they could hold the flanks and *prevent the swift downward running tongues from getting* any wider, that would be enough for the present. [Stewart, 1948, *Fire* (Fiction)]

(80) Resistances to the sociological study of British institutions appear not to have *prevented the Colonial Social Science Research Council from sending* teams of anthropologists into the field. [Birnbaum, 1971, *Toward a Critical Sociology* (Academic)]

(81) Simple filters cannot *stop the near-microscopic zebra mussel larvae from entering* water intake pipes, where they eventually anchor along interior surfaces. [Walker, 1991, *Dreissena Disaster* (Magazine)]

In (79), (80) and (81), each noun phrase contains an average of 5 to 6 words within it. One can observe that there is a lot of detailed information provided through those long noun phrases in order to demonstrate explicitness within the complementation clauses. In (79), *the tongues* are described with three different adjectives, *swift downward running*. In (80), the noun phrase is packed with five different head nouns, *Colonial Social Science Research Council* which is to inform more like those examples provided by Biber (2003: 174) in Section 2.5.1 (Chapter 2). This applies the same for (81) which the noun, *larvae* is described with two other head nouns, *zebra mussel*.

As there are no signs of long noun phrases in the 19th century samples until the late 20th century, this gives an indication that Standard American English has changed over the past decades as this study focuses on 19th to 21st century American English. The current trend of using long noun phrases is a sign of linguistic change in progress. This trend is strongly motivated by the linguistic pattern, densification of content as suggested by Leech et al. (2009: 249-252). The process of *densification of content* means compacting more specific information into noun phrases that contain a sequence of nouns (Leech et al., 2009: 249). Biber et al. (1999: 590) mention that noun sequences are 'extremely dense packaging of referential information'. When those long noun phrases are used widely in the *prevent* and *stop* complementation clauses, they give an implication that there is a tendency towards a more abstract, conceptual and dense writing style in the 21st century. According to Biber and Clark (2002: 63), there is an extremely rapid rate of change in the past 100 years for nouns usage. These changes have been reported in informational written registers like newspaper reportage. Leech et al. (2009: 252) find that greater specialization of topics and audiences are expected at present day and in the future. This

writing style is likely to address to a restricted specialist readership rather than a lay public. An example is specialist readership is tuned into humanistic and philosophical thoughts for social science topics (cf. Leech and Smith, 2005: 89). As we are living in the 21st century internet-era globalized world, there are increasingly more readers who wish to obtain specific and detailed topics for their readings in order to gain more knowledge and have a better understanding about happenings around the globe. Hence, the process of densification of content is considered useful to provide such readings through writings.

Linking back to the process of language change, this particular pattern, densification of content has strong influences in the *prevent* and *stop* complementation clauses. Barber (1964: 142-143) argues that the process of language change is a sign of syntactic and stylistic development. Thus, this development may give a distinctive flavour to Standard American English. As English develops further, there is a tendency for noun phrases to grow longer which comes under the strong influences of technological development, increasing social complexity and the growth of information needs in our modern world. These longer noun phrases are used to promote more efficient and economical communication amongst English speakers (Biber and Clark, 2002: 63-64).

My analysis results have shown the degree of densification of content based on contexts only. However, the compactness of meaning can also be achieved by morphology and processes of word-formation (Leech et al., 2009: 250). They give an example by considering the word, *densification* itself. It is awkward to use more than one word to express the meaning of

densification which is ‘to become more semantically compact’. Thus, the single word, densification sounds much comfortable. Three elements of meanings were condensed into one single abstract noun – ‘dense’ + ‘[dens]-ify’ + ‘[densif]-ication’ to illustrate the meaning of ‘the process of causing something to become denser’. Leech et al. (2009: 251) find that there is a high usage rate of abstract nouns with suffixes such as *-tion, -sion, -ment, -ness, -ity, -ance, -ncy, -acy, -ism, -ship* and *-archy* in American English based on the Freiburg-LOB Corpus of British English (FLOB). The growth rate of abstract nouns is fast especially in academic genre. This implies that there is a tendency towards more abstract and dense writing (Leech et al., 2009: 251).

Leech et al. (2009: 252) say that ‘one could argue that the current trend in written English is to pack ever more information into a given length of text and then to ‘sell’ this fairly heavy intellectual diet in a somewhat more informal (colloquial) style than used to be the case’. Still, it is difficult to argue that this densification of content phenomenon is moving towards speech-like styles and is influenced by the linguistic factor, colloquialization. Sociolinguistic factors such as culture, lifestyle and community also play important roles in influencing the English language to change.

To conclude, the analysis of the length of noun phrases shows a tendency for the noun phrases to increase in length towards the 21$^{st}$ century – a pattern which coincides with the literature (cf. Biber, 2003: 174-177). My results do not demonstrate these changes very clearly and also do not show the speed of the process of language change. This is because of the simplistic methodology which I have chosen to conduct my study. Despite this, the observed changes still indicate that

the process of language change is taking place based on the different stages of increasing and decreasing trends. The reason as why the *prevent* and *stop* complementation clauses favour the with *from* variant in Standard American English is still puzzling and needs more investigation especially in the 21st century data. Other social cultural factors may also play a role in determining its changes and time shall answer these changes.

## 5.4 Summary

This chapter discussed the results of three analyses: (1) the overall frequency per decade for the *prevent* and *stop* complementation clauses, (2) the proportional values comparison of the with *from* and without *from* variants, and (3) the length of noun phrases based on the average number of words within the noun phrases. All the analyses demonstrated evidence of favouring of the with *from* variant for both *prevent* and *stop* complementation clauses in general since the 19th century until the beginning of the 21st century. The *stop* complementation clauses showed slower development in comparison to the *prevent* complementation clauses based on lower frequencies.

The diachronic results showed that *prevent NP from -ing* had a more consistent development trend than *stop NP from -ing*. As the pace of development has almost reached a halt as predicted by Mair (2002: 112), American English will retain only one trend as dominant which is the with *from* variant. Both *prevent NP -ing* and *stop NP -ing* showed decreasing trends towards the 21st century which supported Mair's prediction. When comparing the length of noun phrases for both verbs, the longer the noun phrases, the more these phrases favoured the use of the preposition or complementiser *from*. The observed results supported Rohdenburg's Complexity Principle

(1996: 151) – the more explicit samples tend to favour the cognitively more complex environments. The pattern of linguistic change in progress, densification of content has highly influenced the length of noun phrases within the complementation clauses. Longer noun phrases were used to provide more precise and concise information.

The next chapter will analyse the *-ing* forms in the *prevent* and *stop* complementation clauses through diachronic distinctive collexeme analysis to find out about the preferences for the different semantic domains within the *-ing* forms.

# CHAPTER 6

# DIACHRONIC DISTINCTIVE COLLEXEME ANALYSIS

## 6.1 Overview

This chapter examines the diachronic distinctive collexeme analysis in the complementation clauses of the *-ing* forms of *prevent* and *stop* from COHA. The main aim of this analysis is to measure the degree of attraction or repulsion of the *-ing* forms to the complementation clauses across a period of 200 years and to explore possible preferences changes in their semantic domains. My hypothesis is a shift in the semantic domains from the analysis signifies linguistic change in progress.

First, an overview of collostructional analysis is outlined (Section 6.2). This is followed by a detailed discussion of diachronic distinctive collexeme analysis (Section 6.3). Next section (6.4) describes the methodology which is used generally in any collostructional analysis. This is followed by a related case study from Hilpert (2006) in Section 6.5. The methods of analysis which I have taken to conduct the diachronic distinctive collexeme analysis are discussed in Section 6.6. A set of results from the analysis are presented for *prevent* and *stop* when used with and without *from* based on different time periods to demonstrate the differences or similarities of the *-ing* forms between those two variants (Section 6.7). These results are compared between both verbs as they have shown similar trends in the overall frequency distribution which is the

favouring of the with *from* variant. The final part of this chapter provides a summary and recaps the principal findings (Section 6.8).

## 6.2 Collostructional analysis

In recent years, corpus linguists start to focus their research more on syntactic phenomena (cf. Gries and Stefanowitsch, 2004a and Hilpert, 2006) due to the increased availability of parts of speech tagging corpora such as the BNC, Brown Family of Corpus and ICE-GB. They begin by investigating the interaction between lexemes and grammatical patterns. Lexicon and grammar were traditionally regarded as completely different phenomena – lexicon consists of specific lexical items and grammar consists of abstract syntactic rules. Later on, both lexicon and grammar were treated as pairs. A study done by Hunston and Francis (2000) assumes that lexicon and grammar are not qualitatively different. They identify and describe the close relationship found between lexemes and grammatical patterns. Finally, they come out with some interesting accounts of grammar than the normal explanations in descriptive grammar (cf. Francis et al., 1996 and 1998). The *Cobuild English Grammar* (Sinclair, 1990) is the outcome of those studies.

According to Hoey (2005: 2), lexis is a very complex but systematic structure and the outcome of this lexical structure is grammar. The interaction of lexis with syntax, phonology and semantic can be achieved through the naturalness of language. A key factor of this naturalness is collocation. Certain corpus linguists, especially Sinclair (1991) and Stubbs (1996), agree that all lexical items have collocations. Hoey (2005: 2) defines a *collocation* as the property of language

where two or more words seem to appear frequently in each other's company (e.g. *inevitable* + *consequence*). *The American Heritage Dictionary of the English Language* (1992: 373) defines a *collocation* as an arrangement of juxtaposition, especially of linguistic elements such as words. In simpler terms, a collocation is the recurrent combination of words.

Sinclair (1991: 170) mentions that collocations can be important in the lexical structure of the language because they are frequently repeated. The study of collocations has developed slowly with the arrival of large corpora and has expanded their meanings within the measures of statistical significance. The statistical definition of a *collocation* is the relationship of a lexical item has with other items that appear with greater than random probability in its context (Hoey, 1991: 6-7). The random probability can be measured by using statistical tests such as MI (mutual information), *t* or *z* scores. Sinclair (1991: 170) also states about three other useful technical terms which are related to collocations – node, collocates and span. A *node* is a word that is being investigated and is always shown at the centre of the concordance lines (Hoey, 2005: 4-5). A *collocate* is a word which occurs in close proximity to a word under investigation (Sinclair, 1991: 170). A *span* is the measurement (in words) of the co-text of a word selected for a study (Sinclair, 1991: 175). A span of -4 and +4 means that four words on either side of the node will be taken as relevant verbal environment An example, the word *dog* is a node in this sentence – 'A cute tiny black *dog* sleeps under the tree'. There is a span of four to the left of the node and four to the right of the node. The words *a, cute, tiny, black, sleeps, under, the* and *tree* are all considered as collocates of this node, *dog*.

Collocational analysis which involves node word's collocates has generated many important insights but tends to ignore the grammatical structures which involve a search word (node) and its respective collocates. It always assumes that high raw frequencies collocates will sort out the relevant results from the accidental ones. This approach is imprecise and could not distinguish some abstract constructions which contain specific lexemes. Thus, Stefanowitsch and Gries (2003: 214) recommend a type of collocational analysis which is more sensitive to specific constructions at various linguistic structure levels – collostructional analysis. *Collostructional analysis* is defined as an analysis which starts with a particular construction and investigates the lexemes which are strongly attracted or repelled by that construction (Stefanowitsch and Gries, 2003: 214). This analysis emphasises on the meaning of constructions with preferences for certain lexical elements which are based on their contexts.

The methods of analysis for collostructional analysis are closely related to Construction Grammar. Hence, the term, construction is an important key in this methodology. A *construction* is defined as a form-meaning correspondence that exists independently of particular verbs (Goldberg, 1995: 1). Traditional grammarians like Chomsky (cf. 1957) argue that constructions have their own meanings which do not depend on the words in the sentences and engage their central role along with construction-specific rules. This notion however has been criticised because it was assumed to be the only way to obtain generalisations across patterns. Over time, syntactic constructions which arise from those principles lead to a new focus on idiosyncratic properties of particular sentence patterns (cf. Levin, 1993). These idiosyncratic properties have always credit lexical items when recognising pattern-specific properties.

The version of Construction Grammar which was used to develop the collostructional analysis methodology was developed by Lakoff (1987) and Goldberg (1995). The theory views constructions as the basic units of linguistics organisation and defines a *construction* as a pairing of form with meaning or use such that some aspects of the form or some aspects of the meaning or use are not strictly predictable from the component parts or from other constructions already established to exist in the language (Goldberg, 1996: 68). In simpler terms, a construction is a linguistic expression that is associated directly with a particular meaning or function which cannot be imitated. The linguistic system is observed as a continuum of consecutively abstract constructions which starts with words to expressions and constructions.

The term, *collostructional analysis* derives from a blend of collocation and construction (Gries and Stefanowitsch, 2004a: 100). This methodology of association measures the co-occurrences of words (collocations) that are strongly attracted or repelled to any syntactic patterns (constructions). Some important and useful terminology in this application is discussed below (Stefanowitsch and Gries, 2003: 215):

(i) Collexeme – a lexeme that is attracted to a particular construction (e.g. A *disaster* is waiting to happen.)

(ii) Collostruct – a construction that is associated with a particular lexeme (e.g. A disaster *is waiting to happen*.)

(iii) Collostruction – a combination of a collexeme and a collostruct (e.g. *A disaster is waiting to happen* – [NP be waiting to happen])

According to Gries (to appear), there are three different types of collostructional analysis application which have been used in a variety of fields and languages to yield systematic quantitative results including complementation patterns, constructional senses, verb-specific syntactic priming effects, syntactic alternations of a variety of constructions and distributional patterns in second language learning. They are as follows:

(i) Collexeme analysis – quantifies the degree of attraction or repulsion of words which are typically verbs to a syntactically defined slot in a construction (cf. Stefanowitsch and Gries, 2003, e.g. how much does *give* prefer to occur in a ditransitive construction)

(ii) Distinctive collexeme analysis – quantifies which words (also typically verbs) are attracted to or repelled by one of several constructions (cf. Gries and Stefanowitsch, 2004a, e.g how much does *give* prefer to occur in a ditransitive construction as opposed to a prepositional dative construction)

(iii) Covarying collexeme analysis – identifies preferred and dispreferred pairs in two slots of one construction (cf. Gries and Stefanowitsch, 2004b, e.g. how much do *tricked* and *marrying* prefer to occur in '*He tricked her into marrying him'*)

As this study aims to find out which *-ing* forms have stronger attraction or repulsion to each complementation clauses, I chose to use the distinctive collexeme analysis application. This is because I have to contrast the *-ing* forms between two comparable variants (with *from* and without *from*). As my data consists of samples over a period of 200 years, I will follow Hilpert's

methodology (2006) to conduct the distinctive collexeme analysis diachronically of which I will divide the data into different time periods.

## 6.3 Diachronic distinctive collexeme analysis

According to Gries and Stefanowitsch (2004a: 100), in much corpus linguistic research, the linguistic context of a particular lexeme holds vital clues to its semantic and syntactic properties. Thus, analysing the context based on lexemes is essential. These lexemes are investigated based on syntactic and semantic structures which allow interactions between single words and grammatical constructions. One recommended methodology for investigation which is recommended by Stefanowitsch and Gries (2003) is collostructional analysis. As proposed by Gries (to appear), one of the applications of collostructional analysis is distinctive collexeme analysis. The distinctive collexeme analysis is used for the observation of different degrees of attraction or repulsion of lexemes in two comparable grammaticalised constructions (cf. Gries and Stefanowitsch, 2004a). Hilpert (2006) then applies this distinctive collexeme analysis to historical data which makes use of diachronic corpora and called it diachronic distinctive collexeme analysis.

The main purpose of a diachronic distinctive collexeme analysis is to systematically track the changes of collocational preferences to certain constructions in different time periods. According to Goldberg's sense (1995) in Hilpert (to appear), collocational preferences are referred as the forms and meanings of grammaticalised constructions which are typically displayed by changes in phonology and morpho-syntax or semantic shifting towards more abstract meanings. Certain

verbs (collocates) in some constructions tend to appear more often than in other constructions which result to a more preferred semantic categorisation. The idea behind this analysis is a shift in the collocational patterns may indicate a shift in their semantic domains (Hilpert, 2006: 243). Certain collocates can become more or less favoured over time. The most frequently occurring collocates during this century could vary with those that occurred in the next century due to the process of language change such as grammaticalization, or expansion and reduction of meanings in certain collocates.

Hilpert (to appear) claims that the quantitative analysis of a shift in collocational preferences will provide deeper insights to the semantic development of grammaticalised constructions across different periods of time. A diachronic distinctive collexeme analysis will display the development of collocates to future meanings and is applicable to wider selection of contexts. A synchronic distinctive collexeme analysis does not take into account the overall changes and frequency of certain collocates as it highlights only the differences of the analysed constructions and not the characteristics of them. Collocates may change in frequency over time as well as their characteristics and behaviours. Hence, a diachronic analysis may enable the characterisation of those behaviours and preferences.

This diachronic analysis could be applied to different comparison studies such as the comparison between children and adults on distributional patterns in language learning across time. Infants less than a year old can observe statistical co-occurrence patterns in their own baby-talk language. Their linguistic representation and processing exhibit language change in frequency and

conditional-probability effects. Thus, the measurement of statistical associations between different linguistic elements can tell us about the development of infants' linguistic features over different time periods. For adults, it may be possible to predict their language learning experiences through the results of grammatical patterns in the analysis and to improve on them. Language learning is a time-taking process and thus, the diachronic distinctive collexeme analysis is suitable to track those changes.

Another application of the diachronic distinctive collexeme analysis is to study how certain lexemes have developed over time. An example is the English auxiliary *can* which is used to denote mental ability (Heine, 1993: 90). Consequently, *can* expands semantically over time and takes the infinitive complements that express the actions of sentient human beings such as *say* or *agree*. The selectional restrictions gradually loosen and *can* occurs with a wider set of collocates. These new lexical elements in those constructions thus signify the on-going semantic change. The expansion of collocational preferences put forward to deeper understanding of the development of semantic and syntactic changes across time. Therefore, the results from the analysis will enable us to explore more about the expansion of *can* in various constructions. The methodology of how a diachronic distinctive collexeme analysis is conducted will be explained in the next section.

## 6.4 General methodology of distinctive collexeme analysis

The methodology for distinctive collexeme analysis was introduced by Gries and Stefanowitsch (2004a) to identify those lexemes that exhibit a strong preference for two semantically or

functionally near-equivalent constructions. This methodology was an extension from the work of distinctive collocate analysis by Church et al. (1991) which uses a variant of *t*-test to measure the dissimilarity of semantically similar words on the basis of their lexical collocates. Hilpert (2006) extended the work of Gries and Stefanowitsch (2004a) into diachronic distinctive collexeme analysis which uses similar methodology with distinctive collexeme analysis. The only difference is that the diachronic analysis uses lexemes from diachronic corpora and analyses them based on different time periods.

According to Stefanowitsch (2005), the distinctive collexeme analysis uses a lexeme which is attracted to either one of the two comparable constructions in the analysis. Some examples of constructions that can be used as comparison are ditransitive, prepositional dative constructions, argument structured and verb-specific structures. That lexeme which is being analysed is referred as a distinctive collexeme in those respective constructions. There are two types of frequency which will be used in this analysis – observed frequency and expected frequency. The observed frequency is the frequency which is obtained from the raw frequencies of those investigated lexemes. The expected frequency is the frequency which is generated by the statistical software when the observed frequency is entered in it during the test.

The observed frequency of the lexeme in both constructions is recorded in a two-by-two co-occurrence table (refer Table 14 below). These frequencies are then analysed in the Fisher-Yates Exact Test (cf. Pedersen, 1996) to measure the degree of attraction or repulsion of that particular lexeme in those two comparable constructions. The strength of association is referred as

distinctiveness (Stefanowitsch, 2005). According to Howell (2010: 147), R. A. Fisher (a British biostatistician) introduced *Fisher-Yates Exact Test* in 1934. This test uses a 2 x 2 contingency table by extending the table to include the marginal row and column totals of frequencies as well as the overall total. The Fisher-Yates exact probability calculation works out the probability of getting obtained data. The formula for Fisher-Yates Exact Test is always obtained using statistical software.

Firstly, two comparable constructions which will be investigated in the analysis are determined. All the collexemes of those two constructions are extracted out from the corpora. The time periods for the collexemes are decided and they are listed according to their respective periods. Next, two comparable frequency lists of the potential distinctive collexeme are created. The total frequencies of those two constructions are noted in order to calculate the observed frequency of the listed collexemes. These frequencies are inserted in a two-by-two co-occurrence table for each lexeme. There should be four different frequencies in each lexeme's table as shown in Table 14 below:

Table 14: An example of a two-by-two co-occurrence table (Stefanowitsch, 2005)

| | **Construction 1** | **Construction 2** | **Row totals** |
|---|---|---|---|
| Lexeme Y | A | B | Totals of A and B |
| Other lexemes | C | D | Totals of C and D |
| Column totals | Totals of A and C | Totals of B and D | Grand total of A – D |

The representativeness of the letters is as follows:

(i) A – the observed frequency of lexeme Y in Construction 1

(ii) B – the observed frequency of lexeme Y in Construction 2

(iii) C – the observed frequency of other lexemes beside lexeme Y in Construction 1

(iv) D – the observed frequency of other lexemes beside lexeme Y in Construction 2

As the tables for each lexeme are set up, they are submitted in the Fisher-Yates Exact Test by using any statistics software. An example of a frequently used statistical software is R. R is an open source programming language which is available freely online (refer http://cran.at.r-project.org). The results for each lexeme are sorted out according to its degree of distinctiveness which is the association of strength between lexemes and constructions. A large value indicates strong attraction between the lexeme and its constructions while a small value indicates repulsion between the lexemes and its constructions. The results are finally interpreted in order to find out whether is there any semantic preference in each time period for each variant and do the preferences change over time.

Based on the general methodology discussed, the diachronic distinctive collexeme analysis can be used to outline the potential of a collostructional approach to meaning construction in different cultures and applications as it denotes a concept which is fundamental to many researches in corpus linguistics. Hence, I will relate this diachronic distinctive collexeme analysis to a case study by Hilpert (2006) which is relevant to my study. As the methodology discussed above is systematic and easy to follow, I will employ it to find out the semantic domain preferences for the *-ing* forms of the *prevent* and *stop* complementation clauses when used with *from* and without *from* across three centuries.

## 6.5 Case study

The distinctive collexeme analysis methodology can be applied on either synchronic or diachronic corpus data. Previously, Gries and Stefanowitsch (2004a: 113-115) have studied on two English constructions (*will* and *be going to*) by using the data collected from the International Corpus of English (ICE-GB). The results have shown the differences which lie in the preferred verbal collocates of those two constructions. *Be going to* decides on more dynamic and specific actions or events than *will*. Greater certainty is needed for more specific actions and they also require more efforts which are associated with *be going to*. Nevertheless, these differences do not directly affect the raw frequencies of the respective collocates in Hilpert's study (2006).

Hilpert (2006: 243-247) first studies the preferred verbs in two comparable English constructions (*will* and *be going to*) based on the British National Corpus (BNC) via the synchronic methodology which was used by Gries and Stefanowitsch (2004a) as mentioned above. According to Hilpert (2006: 244), both constructions are used to refer to future events but several differences between them have been mentioned (Binnick, 1971, Wekker, 1976, Close, 1977, Haegeman, 1989 and Berglund, 1997 – all in Hilpert, 2006: 244). Leech (1992) lists the top ten most frequent verbs in those two constructions based on the data from the BNC. They are as follows:

Table 15: Top 10 most frequent verbs with *will* and *going to* in the BNC (Hilpert, 2006: 244)

| ***Will*** | **Tokens** | ***Be going to*** | **Tokens** |
|---|---|---|---|
| *Be* | 41947 | *be* | 4756 |
| *have* | 5906 | *do* | 1907 |
| *Take* | 4150 | *get* | 1403 |
| *make* | 3182 | *have* | 983 |
| *Do* | 3039 | *take* | 647 |
| *Go* | 2821 | *say* | 643 |
| *come* | 2732 | *make* | 631 |
| *Give* | 2543 | *go* | 616 |
| *continue* | 2477 | *happen* | 552 |
| *Find* | 2465 | *tell* | 434 |

Table 16: Top 10 distinctive collexemes of *will* and *going to* in the BNC (Hilpert, 2006: 245)

| ***Will*** | **CollStr** | ***Be going to*** | **CollStr** |
|---|---|---|---|
| *continue* | 83.57 | *do* | infinitive |
| *Be* | 74.17 | *get* | infinitive |
| *provide* | 61.39 | *say* | 195.36 |
| *include* | 56.35 | *happen* | 135.34 |
| *remain* | 44.76 | *ask* | 87.20 |
| *receive* | 42.50 | *die* | 78.72 |
| *become* | 41.15 | *put* | 74.96 |
| *depend* | 39.41 | *tell* | 58.85 |
| *enable* | 37.72 | *marry* | 53.99 |
| *require* | 36.58 | *let* | 42.95 |

From Table 15 above, the results show some collocational overlapping. Verbs like *be*, *do*, *have* and *go* are the most frequent appearing verbs which have a wide range of semantic meanings. The distinctive collexeme analysis will move away from those frequently appearing verbs and determine the asymmetries in frequencies of the co-occurring lexical verbs. This methodology ranks those more significantly occurring lexical verbs in one construction than another to a distinctive level. A mathematical analysis of the Fisher-Yates Exact test is performed based on

those verbs. The results in Table 16 above demonstrate that there are more non-agentive or low in transitivity verbs which are more attracted to the *will* constructions such as *continue, include, remain* and *depend*. For *be going to*, the verbs that are highly attracted are *do, say, put* or *marry* which lie within the agentive and high in transitivity semantic preference.

This analysis can be applied with the same methodology across different time periods to study the development of different grammatical constructions precisely and to engage into theoretical discussions. Hilpert (2006: 247-250) uses the diachronic corpora – The Penn-Helsinki Parsed Corpus of Early Modern English (PPCEME) and the Corpus of Late Modern English Texts (CLMET) to compare the collocations in the *shall* constructions at different historical stages. He first extracts all the instances of *shall* with its orthographical variants and inflected alternatives. Orthographical variants are standardised such as *fynde* to *find*. Then, he categorises them into three different time periods. The two corpora are divided into three 140-year time periods for the purpose of this study. A frequency list of infinitive complements of the auxiliary is identified.

Table 17: Top 10 most frequent verbs with *shall* over three time periods (Hilpert, 2006: 248)

| **1500 – 1640** | | **1640 – 1780** | | **1780 – 1920** | |
|---|---|---|---|---|---|
| **Verb** | **Tokens** | **Verb** | **Tokens** | **Verb** | **Tokens** |
| *Be* | 736 | *be* | 557 | *be* | 1074 |
| *have* | 291 | *have* | 234 | *have* | 527 |
| *Find* | 133 | *find* | 107 | *see* | 239 |
| *See* | 131 | *see* | 75 | *go* | 195 |
| *come* | 120 | *make* | 69 | *do* | 176 |
| *Do* | 117 | *think* | 57 | *find* | 116 |
| *make* | 94 | *take* | 52 | *take* | 95 |
| *Take* | 92 | *endeavour* | 52 | *make* | 89 |
| *hear* | 73 | *do* | 51 | *say* | 87 |
| *know* | 69 | *give* | 46 | *get* | 82 |

Table 18: Top 15 distinctive collexemes of *shall* over three time periods (Hilpert, 2006: 249)

| **1500 – 1640** | | | **1640 – 1780** | | | **1780 – 1920** | | |
|---|---|---|---|---|---|---|---|---|
| **Verb** | **N** | **CollStr** | **Verb** | **N** | **CollStr** | **Verb** | **N** | **CollStr** |
| *understand* | 48 | 15.48 | *endeavour* | 52 | 16.36 | *forget* | 81 | 17.01 |
| *Come* | 120 | 10.32 | *discover* | 17 | 7.86 | *go* | 194 | 12.91 |
| *forfeit* | 40 | 6.53 | *examine* | 13 | 6.86 | *get* | 81 | 9.46 |
| *perceive* | 19 | 6.52 | *mention* | 18 | 5.90 | *try* | 27 | 6.87 |
| *Bear* | 30 | 6.49 | *suppose* | 14 | 5.67 | *meet* | 53 | 6.36 |
| *appear* | 37 | 5.65 | *confine* | 10 | 5.29 | *feel* | 32 | 5.59 |
| *Serve* | 22 | 5.62 | *direct* | 10 | 5.29 | *have* | 527 | 5.07 |
| *Need* | 28 | 5.48 | *explain* | 12 | 5.14 | *see* | 239 | 4.88 |
| *Eat* | 28 | 5.48 | *think* | 57 | 4.70 | *write* | 45 | 4.11 |
| *Bring* | 40 | 5.28 | *add* | 18 | 4.33 | *return* | 43 | 3.96 |

From the results in Table 17 above, the top 4 verbs that have the most number of occurrence times with *shall* were almost the same (*be, have, find* and *see*). However, the distinctive collexeme methodology abstracts away those items which are the most common in each time period and highlight those that are significantly more frequent than expected. Items are considered significant if they occur with a higher relative frequency in a time period than another.

In this way, the differences between those three time periods can be seen more clearly on how the semantic domain preferences shift over time and what the actual development of those verbs is. Table 18 above shows the results from the diachronic distinctive collexeme analysis.

Based on the results in Table 18, there are large significant mismatches between the raw frequencies and the values of collostructional strength. The strongest collocations are ranked at the top of the table. The most distinctive verb in the first time period, *understand* is nowhere to be seen in the second and third periods. This means that *understand* is more unique in the *shall* constructions than *come* which occurs more times. The results demonstrate a different perspective on the historical data than the raw frequencies obtained. This study concludes by presenting a new set of data which could not be seen through the observation of raw frequencies.

The diachronic collexeme analysis is useful to explore how a given construction changes over time by taking into account the description of semantic changes. It also allows us to explore further into the patterns of linguistic change in progress such as grammaticalization or densification of content. For this study, I have employed the same methodology which is used by Hilpert (2006: 247-250) in order to carry out the test on the *-ing* forms of the *prevent* and *stop* complementation clauses and to figure out the preferred semantic domains for those variants across time.

## 6.6 Methodology

In Chapter 5, the results from the analyses of the diachronic frequency change have shown that the with *from* variant is being favoured in both *prevent* and *stop* complementation clauses. It is observed that the with *from* variant was dominating from the beginning of the 19th century until the early 21st century. However, a question regarding the semantic domain preferences of the -*ing* forms with used with and without *from* was left unanswered. It requires the results from the collostructional analysis. The research question is as follows:

(i) Do certain -*ing* forms show a preference for the with *from* variant or the without *from* variant and are there changes over time?

As there are two similar constructions for comparison, the distinctive collexeme analysis methodology was employed (refer Section 6.3). The classification of verbs which will be used in this analysis is obtained from Biber et al. (1999: 360-364) (refer Section 3.3, Chapter 3).

First, all the -*ing* forms of the *prevent* and *stop* complementation clauses were extracted from the data collection and listed in separate columns as detailed in Section 4.4 (Chapter 4). There were two different lists for two different variants (*prevent / stop NP from -ing* and *prevent / stop NP -ing*). Splitting up the timeline into too many periods may lead to an extremely fragmented overall picture (Stefanowitsch, 2006: 260). Thus, three different time periods were chosen which ranged from approximately 60 to 70 years within the 200 years. They were 1810-1879, 1880-1949 and 1950-2009. This step was similar to the study by Hilpert (2006) in order to achieve an almost equal distribution number of samples in each time period for consistency purpose in the analysis.

Subsequently, the observed frequency of each lexeme (*-ing* forms) for both variants of the *prevent* complementation clauses was counted. The grand total of all the lexemes which occurred in that time period was noted in order to calculate the rest of the observed frequencies needed in the contingency tables. For each lexeme, a two-by-two contingency table was made and the respective frequencies were entered in the table. These steps were repeated for all the lexemes in that particular period. An example of the table is shown below:

Table 19: An example of the frequencies for *being* in the *prevent* complementation clauses

| | **With *from* variant** | **Without *from* variant** | **Totals** |
|---|---|---|---|
| *Being* | 251 | 78 | 329 |
| Other lexemes | 2031 | 290 | 2321 |
| Totals | 2282 | 368 | 2650 |

Next, all the two-by-two contingency tables were arranged in a list as text file. This list was entered in R to obtain the results of the Fisher-Yates exact test. The script which was used in R for this test was written by Stefan Th. Gries (version 28th March 2010) and can be obtained from http://www.linguistics.ucsb.edu/faculty/stgries/teaching/groningen/coll.analysis.r.

The instructions to conduct the test were followed as stated in the script. Then, the results from the test were ordered according to their distinctiveness – the larger the collostructional values, the stronger the attraction of lexemes to the constructions. Similar steps were repeated for the other time periods to obtain a full set of results for the *prevent* complementation clauses. Finally, the top 15 lexemes are classified into different semantic domains. These respective domains may reflect a trend in the process of language change.

All the steps in this process were repeated for the *stop* complementation clauses based on three time periods which is similar to the *prevent* complementation clauses. In the following section (6.7), the results from these analyses are compared to the raw frequencies and they are interpreted.

## 6.7 Results and interpretation

Table 20 below show the results of the top 15 most frequent *-ing* forms which occurred in a comparison of *prevent NP from -ing* with *prevent NP -ing* in three different time periods based on their raw frequencies (N). All the *-ing* forms are also classified into their respective semantic domains proposed by Biber et al. (1999: 360-364).

Based on Table 20 below, it is obvious that a lot of the *-ing* forms overlap for both variants within the three time periods. The verbs, *becoming, getting, going, making, doing, seeing, coming, giving* and *having* appear most frequently in both variants based on the raw frequency (observed frequency). *Being* which tops all the three time periods for both *prevent NP from -ing* and *prevent NP -ing* is not surprising as it is the most frequently used verb in the English language.

Table 20: Top 15 most frequent *-ing* forms in *prevent NP from -ing* and *prevent NP -ing* over three time periods

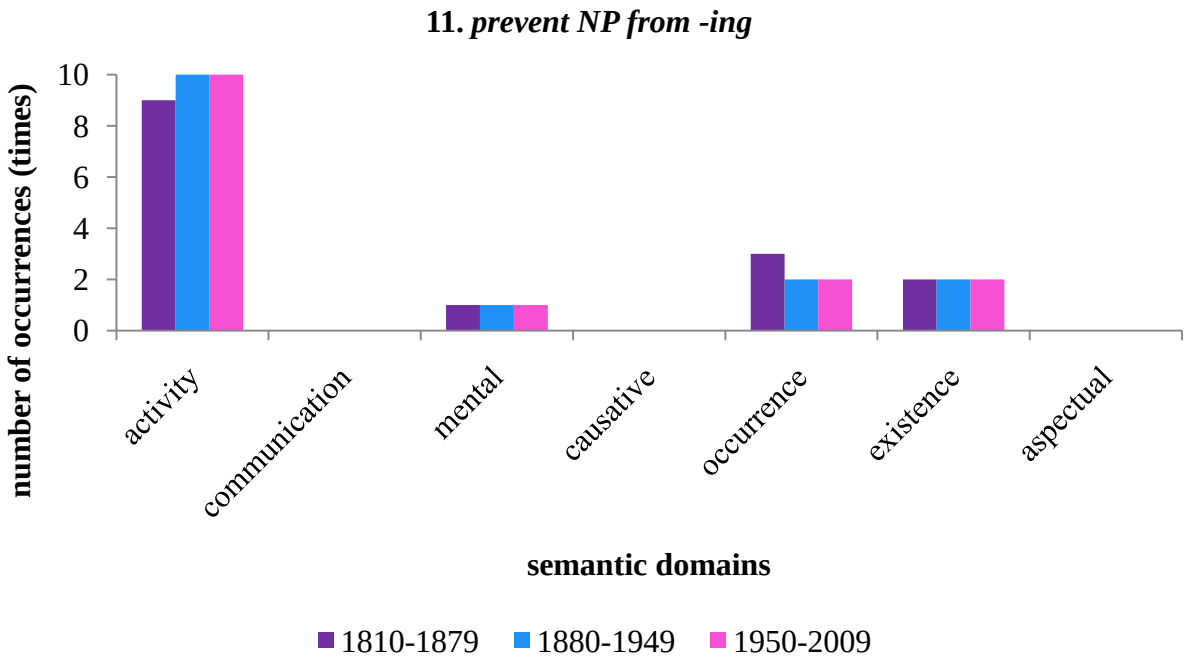
| | ***prevent NP from -ing*** | | | ***prevent NP -ing*** | |
|---|---|---|---|---|---|
| ***-ing* forms** | **Semantic domains** | **N – tokens** | ***-ing* forms** | **Semantic domains** | **N – tokens** |
| **1810 – 1879** | | | | | |
| *being* | existence | 251 | *being* | existence | 78 |
| *becoming* | occurrence | 65 | *going* | activity | 24 |
| *taking* | activity | 63 | *coming* | activity | 12 |
| *going* | activity | 58 | *falling* | occurrence | 10 |
| *making* | activity | 55 | *making* | activity | 8 |
| *doing* | activity | 49 | *seeing* | mental | 8 |
| *seeing* | mental | 46 | *taking* | activity | 8 |
| *falling* | occurrence | 40 | *doing* | activity | 6 |
| *giving* | activity | 36 | *giving* | activity | 6 |
| *getting* | activity | 33 | *having* | existence | 6 |
| *coming* | activity | 23 | *leaving* | activity | 6 |
| *rising* | occurrence | 23 | *becoming* | occurrence | 5 |
| *entering* | activity | 22 | *getting* | activity | 5 |
| *having* | existence | 22 | *receiving* | activity | 5 |
| *obtaining* | activity | 20 | *arising* | occurrence | 4 |
| **1880 – 1949** | | | | | |
| *being* | existence | 302 | *being* | existence | 78 |
| *getting* | activity | 112 | *going* | activity | 22 |
| *becoming* | occurrence | 106 | *getting* | activity | 21 |
| *making* | activity | 98 | *taking* | activity | 18 |
| *doing* | activity | 93 | *coming* | activity | 15 |
| *going* | activity | 92 | *seeing* | mental | 15 |
| *taking* | activity | 91 | *doing* | activity | 14 |
| *falling* | occurrence | 70 | *falling* | occurrence | 11 |
| *coming* | activity | 53 | *becoming* | occurrence | 10 |
| *seeing* | mental | 46 | *making* | activity | 10 |
| *entering* | activity | 41 | *having* | existence | 9 |
| *reaching* | activity | 40 | *using* | activity | 8 |
| *giving* | activity | 38 | *reaching* | activity | 7 |
| *having* | existence | 36 | *leaving* | activity | 6 |
| *carrying* | activity | 35 | *carrying* | activity | 5 |
| **1950 – 2009** | | | | | |
| *being* | existence | 224 | *being* | existence | 37 |
| *getting* | activity | 96 | *coming* | activity | 9 |
| *becoming* | occurrence | 94 | *getting* | activity | 6 |
| *making* | activity | 74 | *doing* | activity | 5 |
| *taking* | activity | 73 | *falling* | occurrence | 5 |
| *doing* | activity | 71 | *going* | activity | 4 |
| *going* | activity | 62 | *making* | activity | 4 |
| *using* | activity | 51 | *turning* | activity | 4 |
| *seeing* | mental | 46 | *following* | activity | 3 |
| *falling* | occurrence | 44 | *seeing* | mental | 3 |
| *reaching* | activity | 44 | *arising* | occurrence | 2 |
| *entering* | activity | 42 | *biting* | activity | 2 |
| *having* | existence | 32 | *ending* | aspectual | 2 |
| *coming* | activity | 30 | *entering* | activity | 2 |
| *moving* | activity | 26 | *escaping* | activity | 2 |

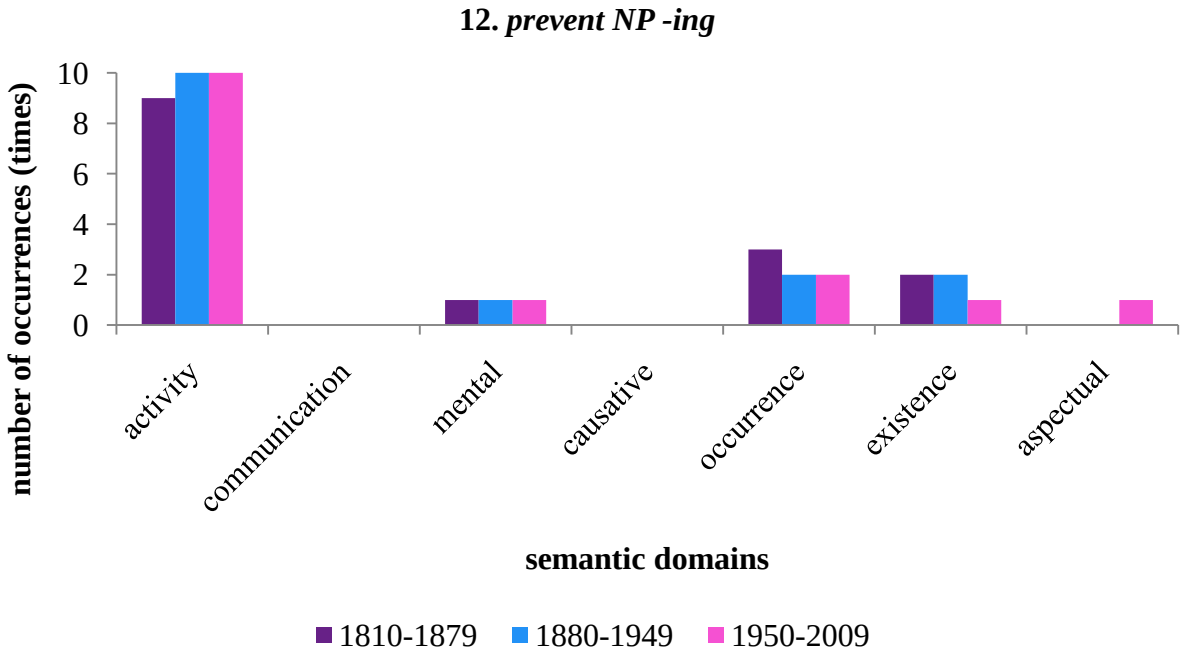
According to the classification of verbs by Biber et al. (1999: 360-364), verbs such as *going, making, coming, moving, carrying, using, getting* and *doing* are classified as *activity* verbs: those which denote actions and events that could be associated with choice and take a subject with the semantic role of agent. *Becoming, arising* and *falling* are examples of *occurrence* verbs: those which report events of non-volitional activities typically physical events. *Seeing* is considered as part of *mental* verbs domain: those which denote activities and state experienced by humans that do not involve physical actions. *Being* and *having* are classified as *existence* verbs: those which report a state which exists between entities. *Ending* is an example for *aspectual* verbs domain: those which characterise the stage of progress of some activities.

For the comparison of the top 15 most frequent *-ing* forms in both *prevent* complementation clauses, there are no occurrences of *causative* verbs (those which indicate a new state of affairs brought by a person or inanimate entity) or *communication* verbs (those which involve communication activities). The verb, *prevent* is classified under causative verbs, hence it is meaningless in a complementation clause to have two causative verbs along side. This might be the reason as why there are no occurrences of causative verbs. However, in the top 30 most frequent *-ing* forms list, there is an occurrence of a communication verb which is *saying*.

Graphical illustrations are used to provide a clearer picture to show the results of the analysis above. Figure 7 (Graphs 11 and 12) below groups the top 15 most frequent *-ing* forms in *prevent NP from -ing* and *prevent NP -ing* into the six semantic domains discussed above based on the number of occurrences (times) across three different time periods.

Figure 15: Top 15 most frequent *-ing* forms in *prevent NP from -ing* and *prevent NP -ing* over three time periods: frequencies in times



Based on Figure 15 (Graphs 11 and 12) above, the same category, activity verbs, dominates from 1810 to 2009. This is not surprising as activity verbs are the most common verbs in English. Based on a study from the Longman Spoken and Written English Corpus (LSWE) by Biber et al. (1999: 382-383), activity verbs are used frequently in transitive, intransitive and copular patterns. The other semantic domains have very low occurrences or are absent. Both bar charts suggest that there are no changes in the semantic domains for both *prevent* complementation clauses in all three time periods. However, it is dangerous to interpret this evidence of changes in semantic domains based on only their raw frequencies.

These results may be interpreted in several ways: (1) there is no change in the semantic domains or the change has become stable over time; or (2) the analysis of the raw frequencies alone does not reveal the development of the complementation clauses; or (3) the analysis does not show which *-ing* forms are strongly attracted or repelled by the constructions. Thus, the diachronic distinctive collexeme analysis which does not depend on the raw frequencies is being conducted for better observation of the development of the *prevent* complementation clauses.

Table 21: Top 15 distinctive *-ing* forms in *prevent NP from -ing* and *prevent NP -ing* over three time periods

| | ***prevent NP from -ing*** | | | | ***prevent NP -ing*** | | |
|---|---|---|---|---|---|---|---|
| ***-ing* forms** | **Semantic domains** | **N – tokens** | **Collstr** | ***-ing* forms** | **Semantic domains** | **N – tokens** | **Collstr** |
| **1810 – 1879** | | | | | | | |
| *obtaining* | activity | 20 | 1.30 | *being* | existence | 78 | 6.84 |
| *becoming* | occurrence | 65 | 1.20 | *going* | activity | 24 | 3.78 |
| *acquiring* | activity | 14 | 0.91 | *arising* | occurrence | 4 | 2.80 |
| *rising* | occurrence | 23 | 0.87 | *coming* | activity | 12 | 2.77 |
| *enjoying* | mental | 13 | 0.84 | *leaving* | activity | 6 | 1.80 |
| *following* | activity | 13 | 0.84 | *facing* | mental | 2 | 1.72 |
| *entering* | activity | 22 | 0.82 | *regaining* | activity | 2 | 1.72 |
| *attending* | occurrence | 12 | 0.78 | *resulting* | causative | 2 | 1.72 |
| *extending* | activity | 12 | 0.78 | *surrendering* | activity | 2 | 1.72 |
| *passing* | activity | 12 | 0.78 | *ascending* | occurrence | 3 | 1.68 |
| *throwing* | activity | 10 | 0.65 | *noticing* | mental | 3 | 1.42 |
| *forming* | activity | 9 | 0.58 | *issuing* | occurrence | 2 | 1.29 |
| *knowing* | mental | 9 | 0.58 | *stopping* | aspectual | 2 | 1.29 |
| *committing* | activity | 8 | 0.52 | *holding* | activity | 3 | 1.07 |
| *slipping* | occurrence | 8 | 0.52 | *loving* | mental | 3 | 1.07 |
| **1880 – 1949** | | | | | | | |
| *becoming* | occurrence | 106 | 1.25 | *being* | existence | 233 | 16.92 |
| *receiving* | causative | 17 | 1.11 | *climbing* | activity | 4 | 2.36 |
| *making* | activity | 98 | 1.02 | *reading* | mental | 4 | 2.04 |
| *holding* | activity | 15 | 0.98 | *riding* | activity | 3 | 2.02 |
| *interfering* | causative | 15 | 0.98 | *seeing* | mental | 15 | 1.76 |
| *turning* | activity | 15 | 0.98 | *injuring* | mental | 2 | 1.71 |
| *working* | activity | 24 | 0.93 | *tending* | existence | 2 | 1.71 |
| *enjoying* | mental | 14 | 0.91 | *wedding* | activity | 2 | 1.71 |
| *feeling* | mental | 14 | 0.91 | *signing* | communication | 3 | 1.67 |
| *starting* | aspectual | 14 | 0.91 | *drifting* | causative | 3 | 1.41 |
| *attending* | occurrence | 13 | 0.85 | *coming* | activity | 15 | 1.36 |
| *keeping* | aspectual | 13 | 0.85 | *creeping* | activity | 2 | 1.28 |
| *realising* | mental | 13 | 0.85 | *happening* | occurrence | 2 | 1.28 |
| *returning* | activity | 12 | 0.78 | *noticing* | mental | 2 | 1.28 |
| *adopting* | mental | 11 | 0.72 | *raising* | occurence | 2 | 1.28 |
| **1950 – 2009** | | | | | | | |
| *becoming* | occurrence | 94 | 1.67 | *being* | existence | 37 | 6.85 |
| *taking* | activity | 73 | 0.75 | *coming* | activity | 9 | 3.53 |
| *using* | activity | 51 | 0.74 | *biting* | activity | 2 | 2.47 |
| *reaching* | activity | 44 | 0.60 | *resulting* | causative | 2 | 2.47 |
| *saying* | communication | 21 | 0.55 | *ending* | aspectual | 2 | 2.01 |
| *giving* | activity | 17 | 0.44 | *arising* | occurrence | 2 | 1.36 |
| *developing* | occurrence | 16 | 0.42 | *turning* | activity | 4 | 1.32 |
| *growing* | occurrence | 16 | 0.42 | *accompanying* | activity | 1 | 1.23 |
| *pulling* | activity | 15 | 0.39 | *affecting* | causative | 1 | 1.23 |
| *having* | existence | 32 | 0.38 | *blackening* | activity | 1 | 1.23 |
| *gaining* | occurrence | 14 | 0.37 | *busing* | activity | 1 | 1.23 |
| *interfering* | causative | 13 | 0.34 | *dallying* | activity | 1 | 1.23 |
| *joining* | activity | 13 | 0.34 | *departing* | activity | 1 | 1.23 |
| *running* | activity | 12 | 0.34 | *draining* | occurrence | 1 | 1.23 |
| *happening* | occurrence | 12 | 0.31 | *drilling* | activity | 1 | 1.23 |

The results from the diachronic distinctive collexeme analysis are shown in Table 21 above. There are considerable differences between the *-ing* forms which are mostly strongly attracted to *prevent NP from -ing* and *prevent NP -ing* in three different time periods. In contrast to the ranking by the raw frequencies, it is not the most frequent verbs that top the results; instead it is the most distinctive verbs (refer Section 6.4). This is displayed in the results, shown by the total number of raw frequencies (N) where some verbs occur only once or twice in comparison to the results in Table 20 above. Nevertheless, these verbs are the most distinctive because they have the strongest attraction to the constructions. The attraction strengths of the *-ing* forms are displayed by their collostructional values (Collstr). The larger the collostructional values, the stronger the *-ing* forms are attracted to the constructions. The smaller the collostructional values, the stronger the *-ing* forms are repelled away from the constructions. The *-ing* forms in *prevent NP -ing* have larger collostructional values than in *prevent NP from -ing* because they occur fewer times in *prevent NP -ing* and thus, the probability of being attracted to *prevent NP -ing* is much higher.

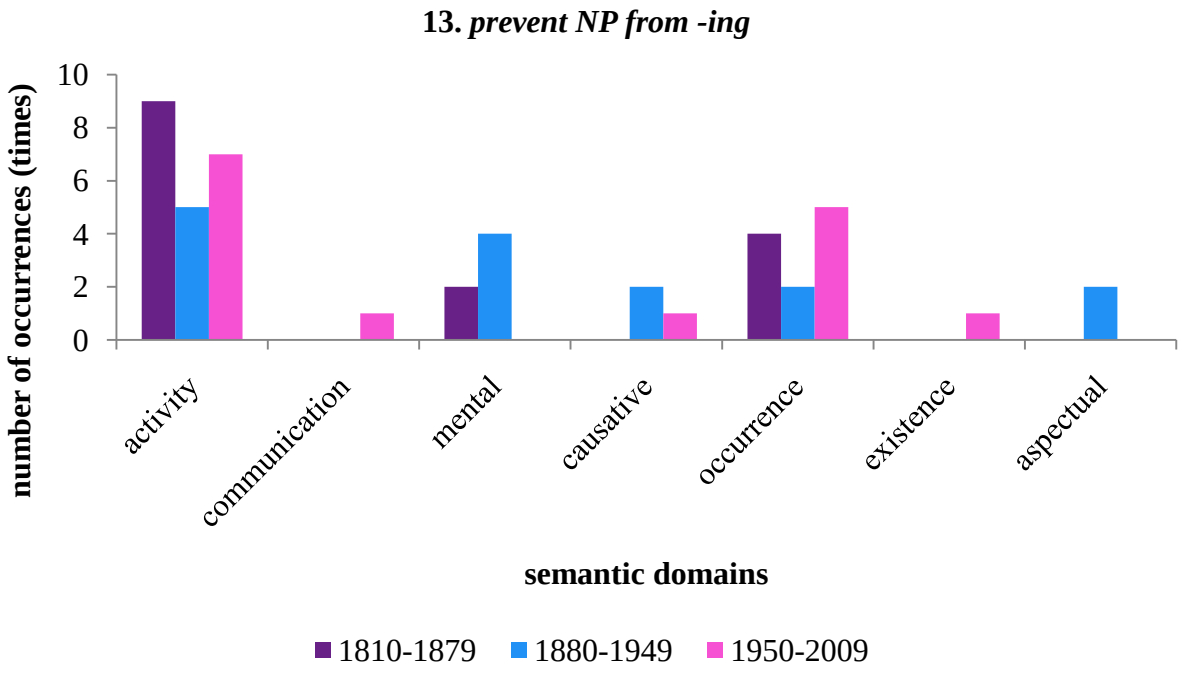
In 1810-1879, activity verbs are observed to have dominated the results shown in Table 21 above based on the values of the collostructional analysis for both types of constructions. In fact, *prevent NP from -ing* have a high number of activity verbs occurrences such as *obtaining, following, entering, extending, passing, throwing* and *forming*. There are four occurrence verbs *becoming, rising, attending* and *slipping* in *prevent NP from -ing* which is the second highest domain after activity verbs. However, there are no occurrences of verbs in the domains such as communication, aspectual, causative and existence. For *prevent NP -ing*, there are more verbs distributing in other semantic domains beside activity verbs. There are a few verbs which are

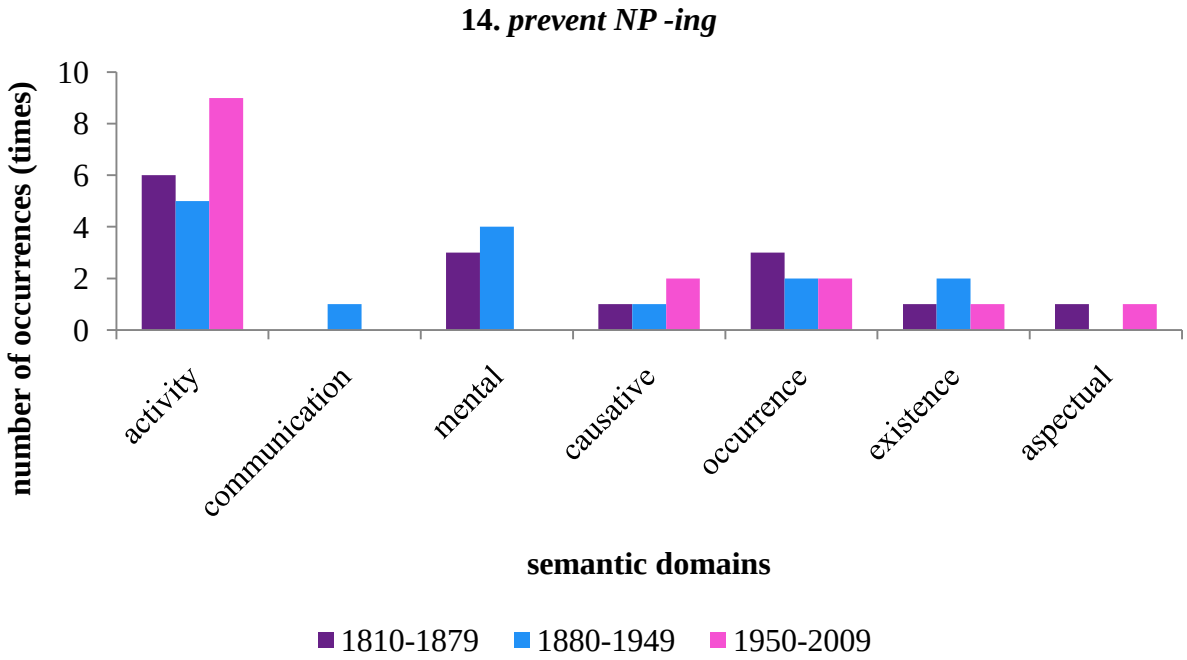
categorised as domains such as mental (*noticing*), occurrence (*arising* and *issuing*), causative (*resulting*) and aspectual (*stopping*).

In 1880-1949, the domination of semantic domains is still the same with those in 1810-1879. However, there are fewer activity verbs which are attracted to both *prevent NP from -ing* and *prevent NP -ing*. Both constructions have only five activity verbs amongst the top 15 distinctive *-ing* forms such as *making, holding, turning, working, climbing, riding* and *wedding*. There are four mental verbs occurrences in both constructions – *enjoying, feeling, realising* and *adopting* for *prevent NP from -ing* and *seeing, injuring, reading* and *noticing* for *prevent NP -ing*. Biber et al. (1999: 365) mention that English speakers usually report on their own opinions and feelings by using mental verbs. Other domains such as causative, aspectual, communication and existence have either low number of verbs being attracted to the constructions or none.

In 1950-2009, it is undoubtedly that activity verbs is the dominating semantic domain especially for *prevent NP -ing*. There is a high number of verb occurrences for *prevent NP -ing* such as *coming, biting, turning, accompanying, blackening, busing* and *departing*. For *prevent NP from -ing*, one interesting observation is a rather high number of occurrences in the occurrence verbs domain. Among the top 15 distinctive ing forms, there are five *-ing* forms in the occurrence verbs domain – *becoming, developing, growing, gaining* and *happening*. There are no distinctive *-ing* forms which are classified as mental verbs in both *prevent* complementation clauses. The rest of the semantic domains have very low number of distinctive verbs.

Figure 16: Top 15 distinctive *-ing* forms in *prevent NP from -ing* and *prevent NP -ing* over three time periods: frequencies in times



In summary as displayed in Figure 16 (Graphs 13 and 14) above, all three time periods (1810-1879, 1880-1949 and 1950-2009) have similar semantic domains preference which is activity verbs. All the *-ing* forms from this domain are attracted strongly to both *prevent NP from -ing* and *prevent NP -ing*. Nevertheless, the period of 1880-1949 shows a decline in the activity verbs domain but increases back in 1950-2009. This domain stabilises in 1950-2009 especially for *prevent NP -ing* which has a high number of verbs attraction in that domain. This is an expected result as activity verbs occur much more commonly than other verbs in English. Based on the study from the Longman Spoken and Written English Corpus (LSWE) by Biber et al. (1999: 365-366), 50% of all the common verbs (verbs that occur at least 50 times per million words) are activity verbs (139 out of 281 common verbs). Hence, my results are similar to the results by Biber et al.

The occurrence verbs domain is the next domain which has a high number of attracting verbs in comparison to the other semantic domains. The communication verbs domain has almost no verbs classified under it. Both of the results are opposite from Biber et al.'s findings (1999: 365-366). Although Biber et al. claim that occurrence verbs are rare; they however are not in my study as *prevent* and *stop* are verbs which are linked closely to happening events. Hence, this might be the reason to why occurrence verbs domain has a high number of *-ing* forms in it. Causative and aspectual verbs domains have only a few verbs classified under them. This is parallel to Biber et al.'s findings which show that causative and aspectual verbs domains are uncommon domains and have low distribution of verbs in them.

This interpretation of semantic domains through the distinctive *-ing* forms across time demonstrate an almost similar perspective with the most frequent *-ing* forms on the development of the *prevent* complementation clauses. Nonetheless, there are still some changes in the distribution of the distinctive *-ing* forms as there are more occurrences in the other domains too. These changes cannot be observed through the raw frequencies of the *-ing* forms. Instead, they can only be interpreted by the collostructional values which were obtained from the distinctive collexeme analysis. An increased number of verbs in the activity verbs domain across the three different time periods indicates that language is changing as English speakers are using more activity verbs in different genres such as conversation, fiction, news and academic writing.

Similar analysis steps are repeated for the *stop* complementation clauses in order to obtain the distinctive *-ing* forms for comparison purpose. Table 22 below show the top 15 most frequent *-ing* forms which appear in *stop NP from -ing* and *stop NP -ing* over three different time periods. Note that the raw frequencies of the *-ing* forms for *stop* are much lower than *prevent* because there are less constructions found in COHA. As many of the *-ing* forms occur only once in the time period of 1810-1879, they might be meaningless for this analysis as they do not provide any information. The development trend of the *-ing* forms cannot be seen clearly too. For the other two time periods which are 1880-1949 and 1950-2009, the *-ing* forms occur more often in comparison to the first time period.

Table 22: Top 15 most frequent *-ing* forms in *stop NP from -ing* and *stop NP -ing* over three time periods

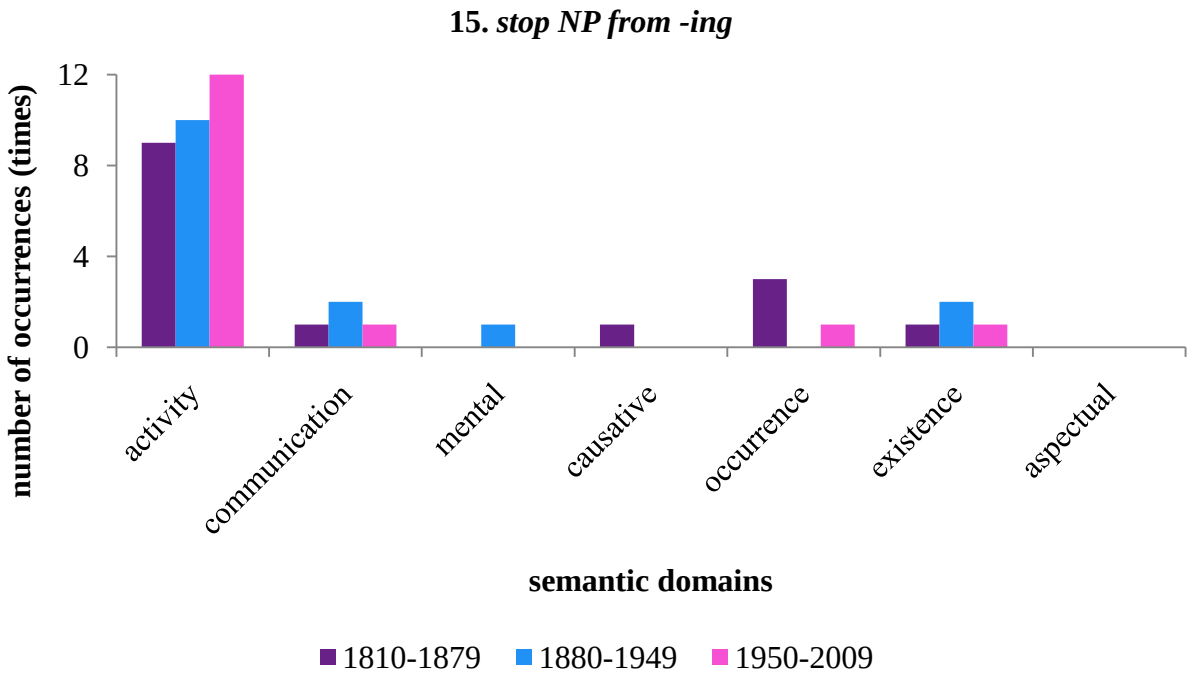
| **1810 – 1879** | | | | | |
|---|---|---|---|---|---|
| ***stop NP from -ing*** | | | ***stop NP -ing*** | | |
| ***-ing* forms** | **Semantic domains** | **N – tokens** | ***-ing* forms** | **Semantic domains** | **N – tokens** |
| *going* | activity | 4 | *carrying* | activity | 2 |
| *taking* | activity | 3 | *taking* | activity | 1 |
| *coming* | activity | 2 | *absolving* | causative | 1 |
| *absolving* | causative | 1 | *blowing* | activity | 1 |
| *attending* | occurrence | 1 | *choking* | activity | 1 |
| *being* | existence | 1 | *denying* | communication | 1 |
| *blowing* | activity | 1 | *fighting* | activity | 1 |
| *choking* | activity | 1 | *following* | occurrence | 1 |
| *denying* | communication | 1 | *gorging* | activity | 1 |
| *dying* | occurrence | 1 | *kindling* | mental | 1 |
| *fighting* | activity | 1 | *reaching* | activity | 1 |
| *flowing* | occurrence | 1 | *rising* | occurrence | 1 |
| *following* | activity | 1 | *attempting* | activity | 1 |
| *giving* | activity | 1 | *catching* | activity | 1 |
| *gorging* | activity | 1 | *chanting* | communication | 1 |
| **1880 – 1949** | | | | | |
| *going* | activity | 27 | *going* | activity | 15 |
| *doing* | activity | 17 | *running* | activity | 5 |
| *being* | existence | 13 | *talking* | communication | 5 |
| *making* | activity | 8 | *containing* | existence | 4 |
| *running* | activity | 8 | *doing* | activity | 4 |
| *coming* | activity | 6 | *flowing* | occurrence | 4 |
| *getting* | activity | 6 | *rocking* | causative | 4 |
| *marrying* | activity | 6 | *screaming* | communication | 4 |
| *trying* | activity | 5 | *being* | existence | 3 |
| *working* | activity | 5 | *coming* | activity | 3 |
| *calling* | communication | 4 | *making* | activity | 3 |
| *having* | existence | 4 | *caring* | mental | 2 |
| *saying* | communication | 4 | *feeling* | mental | 2 |
| *throwing* | activity | 4 | *following* | activity | 2 |
| *bleeding* | mental | 3 | *giving* | activity | 2 |
| **1950 – 2009** | | | | | |
| *doing* | activity | 62 | *going* | activity | 12 |
| *getting* | activity | 44 | *doing* | activity | 9 |
| *going* | activity | 44 | *coming* | activity | 8 |
| *taking* | activity | 43 | *being* | existence | 6 |
| *being* | existence | 38 | *trembling* | mental | 5 |
| *trying* | activity | 21 | *falling* | occurrence | 4 |
| *coming* | activity | 20 | *passing* | activity | 4 |
| *making* | activity | 19 | *playing* | activity | 4 |
| *killing* | activity | 17 | *singing* | communication | 4 |
| *saying* | communication | 16 | *walking* | activity | 4 |
| *falling* | occurrence | 14 | *bleeding* | mental | 3 |
| *leaving* | activity | 14 | *blowing* | activity | 3 |
| *running* | activity | 14 | *dipping* | activity | 3 |
| *using* | activity | 13 | *pacing* | activity | 3 |
| *moving* | activity | 12 | *running* | activity | 3 |

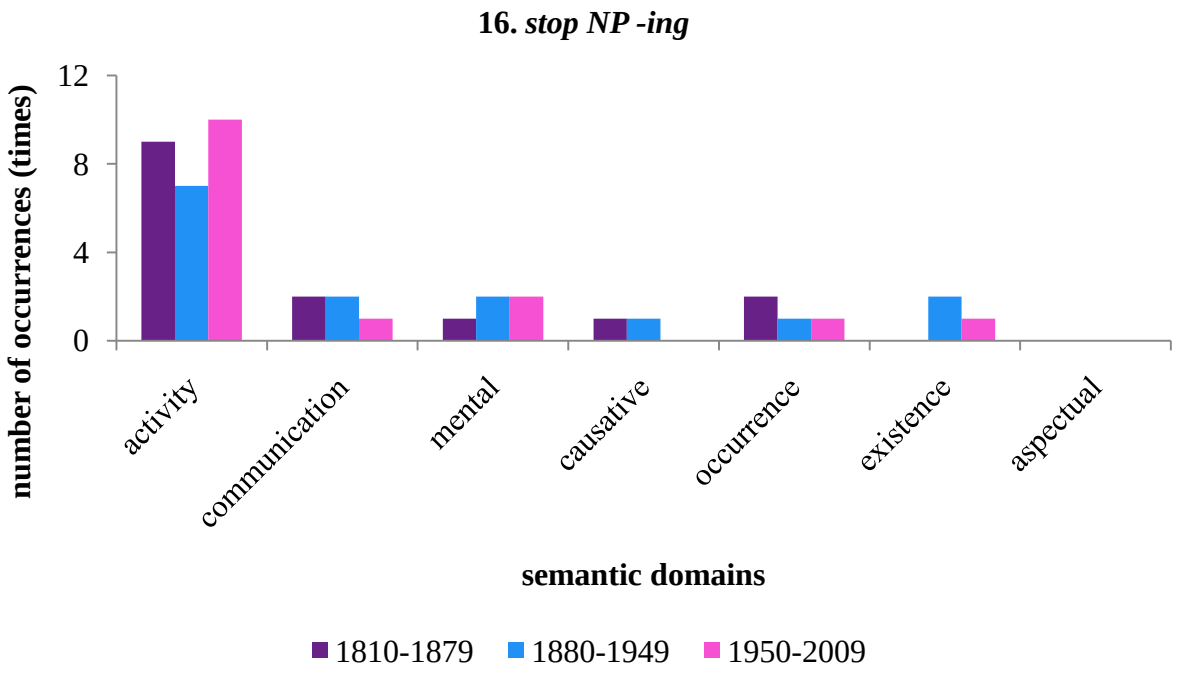
In the results shown in Table 22 above, we see that compared to the *prevent* complementation clauses, the variety of the most frequent -*ing* forms in both variants of the *stop* complementation clauses across time is much more diverse. Only a few verbs (*going, coming* and *doing*) are dominating in all the three time periods for both *stop* variants and they are classified as activity verbs. There is one exception for *stop NP -ing* in 1810-1879 where none of these activity verbs appear in that particular time period. There are many other activity verbs such as *making, running, taking* and *giving* which occur frequently in these *stop* complementation clauses.

*Being* which is classified as an existence verb and occurs in all periods except for 1810-1879 in the *stop NP -ing* pattern. One interesting observation is in all three time periods, there are a few verbs which are classified as communication verbs and this trend differs from the *prevent* complementation clauses. Examples of communication verbs are *singing, saying, talking* and *calling*. In all three time periods, there are a low number of occurrences for causative verbs domain. Aspectual verbs do not appear in the top 15 most frequent -*ing* forms in any of *stop* complementation clauses.

Figure 17 (Graphs 15 and 16) below summarises the number of occurrences (times) of the top 15 most frequent -*ing* forms in both *stop NP from -ing* and *stop NP -ing* into six different semantic domains across all three time periods based on the results from Table 22 above.

Figure 17: Top 15 most frequent *-ing* forms in *stop NP from -ing* and *stop NP -ing* across three time periods: frequencies in times



From Figure 17 (Graphs 15 and 16) above, we can see that activity verbs are the dominating category in all three time periods and especially towards the 21st century. This is not surprising as mentioned earlier because activity verbs are the most common verbs in the English language. Communication verbs domain ranks second after activity verbs domain as it has some occurring verbs in all three periods for both types of *stop* variants. Biber et al. (1999: 365) say that communication is a specialised type of activity; therefore it is an important semantic domain on its own. English speakers distinguish different types of communicative activities and report what others have said or written by using communicative verbs. The other verb semantic domains have low number of occurrences or are completely absent. However, these raw frequencies do not reflect on the development of the *stop* complementation clauses over time because they do not show how strong the *-ing* forms are attracted or repelled to the with *from* and without *from* variants. Thus, the diachronic distinctive collexeme analysis is conducted to obtain the results based on the semantic domains for the *-ing* forms.

Table 23 below displays the top 15 distinctive *-ing* forms which are attracted to *stop NP from -ing* and *stop NP -ing* over three different time periods. From the results, we observe that there have not been many changes in the semantic domains over the three time periods for both *stop* variants. The most frequently occurring verbs do not dominate in the diachronic distinctive collexeme analysis results; instead it is the distinctive verbs (refer Section 6.4). These distinctive verbs have the strongest attraction of the *-ing* forms to the *stop* constructions in the analysis. This can be seen through the comparison of their collostructional values (Collstr) with the total number of occurrences (N). The results show that the strongest *-ing* forms differ within both *stop NP from -ing* and *stop NP -ing*.

Table 23: Top 15 distinctive *-ing* forms in *stop NP from -ing* and *stop NP -ing* over three time periods

| **1810 – 1879** | | | | | | | |
|---|---|---|---|---|---|---|---|
| ***stop NP from -ing*** | | | | ***stop NP -ing*** | | | |
| ***-ing* forms** | **Semantic domains** | **N – tokens** | **Collstr** | ***-ing* forms** | **Semantic domains** | **N – tokens** | **Collstr** |
| *taking* | activity | 3 | 0.57 | *singing* | communication | 2 | 0.94 |
| *going* | activity | 4 | 0.37 | *attempting* | activity | 1 | 0.46 |
| *absolving* | mental | 1 | 0.19 | *beating* | activity | 1 | 0.46 |
| *attending* | occurrence | 1 | 0.19 | *carrying* | activity | 1 | 0.46 |
| *being* | existence | 1 | 0.19 | *catching* | activity | 1 | 0.46 |
| *blowing* | activity | 1 | 0.19 | *chanting* | communication | 1 | 0.46 |
| *choking* | activity | 1 | 0.19 | *dozing* | activity | 1 | 0.46 |
| *denying* | communication | 1 | 0.19 | *entering* | activity | 1 | 0.46 |
| *dying* | occurrence | 1 | 0.19 | *rolling* | activity | 1 | 0.46 |
| *flowing* | occurrence | 1 | 0.19 | *sitting* | activity | 1 | 0.46 |
| *following* | activity | 1 | 0.19 | *splitting* | causative | 1 | 0.46 |
| *gorging* | activity | 1 | 0.19 | *tattling* | communication | 1 | 0.46 |
| *kindling* | communication | 1 | 0.19 | *fighting* | activity | 1 | 0.24 |
| *passing* | activity | 1 | 0.19 | *giving* | activity | 1 | 0.24 |
| *progressing* | aspectual | 1 | 0.19 | – | | | |
| **1880 – 1949** | | | | | | | |
| *getting* | activity | 6 | 1.12 | *containing* | existence | 4 | 1.85 |
| *doing* | activity | 17 | 1.04 | *flowing* | occurrence | 4 | 1.85 |
| *trying* | activity | 5 | 0.93 | *screaming* | communication | 4 | 1.85 |
| *being* | existence | 12 | 0.76 | *rocking* | causative | 4 | 1.29 |
| *calling* | communication | 4 | 0.74 | *talking* | communication | 5 | 1.29 |
| *having* | existence | 4 | 0.74 | *caring* | mental | 2 | 0.92 |
| *throwing* | activity | 4 | 0.74 | *giving* | activity | 2 | 0.92 |
| *marrying* | activity | 6 | 0.62 | *looking* | mental | 2 | 0.92 |
| *bleeding* | mental | 3 | 0.56 | *smuggling* | activity | 2 | 0.92 |
| *fighting* | activity | 3 | 0.56 | *spinning* | activity | 2 | 0.92 |
| *growing* | occurrence | 3 | 0.56 | *feeling* | mental | 2 | 0.56 |
| *putting* | activity | 3 | 0.56 | *selling* | activity | 2 | 0.56 |
| *working* | activity | 5 | 0.49 | *thinking* | mental | 1 | 0.56 |
| *breaking* | activity | 2 | 0.37 | *annoying* | mental | 1 | 0.46 |
| *building* | activity | 2 | 0.37 | *approaching* | activity | 1 | 0.46 |
| **1950 – 2009** | | | | | | | |
| *getting* | activity | 44 | 1.90 | *singing* | communication | 4 | 3.16 |
| *taking* | activity | 43 | 1.37 | *trembling* | mental | 5 | 2.40 |
| *killing* | activity | 17 | 1.32 | *dipping* | activity | 3 | 2.37 |
| *trying* | activity | 21 | 0.99 | *passing* | activity | 4 | 2.10 |
| *telling* | communication | 12 | 0.93 | *pacing* | aspectual | 3 | 1.82 |
| *making* | activity | 19 | 0.86 | *bashing* | communication | 2 | 1.58 |
| *reaching* | activity | 11 | 0.85 | *exploding* | causative | 2 | 1.58 |
| *entering* | activity | 10 | 0.78 | *losing* | activity | 2 | 1.58 |
| *saying* | communication | 15 | 0.63 | *pedalling* | activity | 2 | 1.58 |
| *doing* | activity | 62 | 0.60 | *spilling* | occurrence | 2 | 1.58 |
| *leaving* | activity | 14 | 0.57 | *threatening* | communication | 2 | 1.58 |
| *buying* | activity | 7 | 0.54 | *worrying* | mental | 2 | 1.58 |
| *winning* | activity | 7 | 0.54 | *walking* | activity | 4 | 1.36 |
| *destroying* | causative | 6 | 0.47 | *brushing* | activity | 2 | 1.15 |
| *giving* | activity | 6 | 0.47 | *shivering* | mental | 2 | 1.15 |

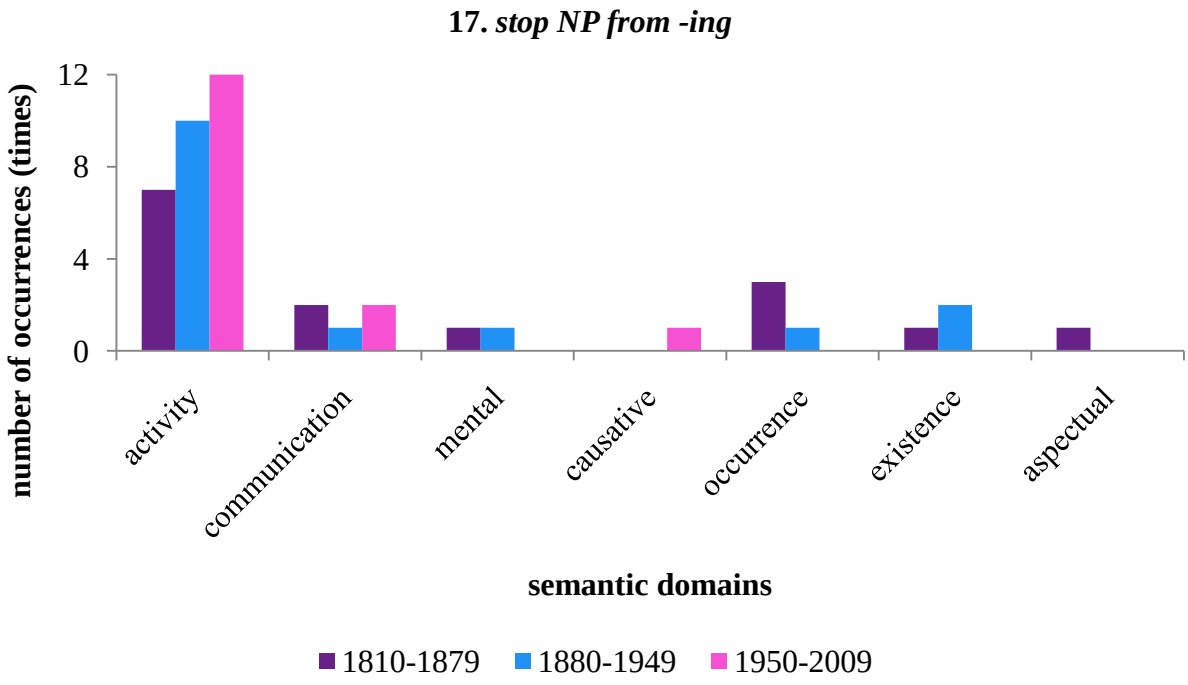
In 1810-1879, based on Table 23 above, most of the distinctive *-ing* forms lie within the activity verbs domain for both *stop NP from -ing* and *stop NP -ing* variants. Examples of activity verbs are *taking, attempting, beating, carrying, catching, blowing, choking, dozing, entering, rolling, sitting, splitting, going, fighting, following* and *gorging.* These collostructional values are low for both variants. Most of the distinctive *-ing* forms occur in this list only by chance as their raw frequencies are very small. This does not reveal much information about the trend and it is pointless for any comparison. There are some communication verbs (*singing, chanting* and *denying*) which are strongly attracted to both *stop NP from -ing* and *stop NP -ing*. The other semantic domains such as occurrence, aspectual and existence verbs have low number of *-ing* forms occurring in their domains for *stop NP from -ing*. However, these three semantic domains do not have any occurrences for *stop NP -ing*. A note to be taken is *stop NP -ing* has only 14 distinctive *-ing* forms as there are not enough verbs to analyse in that particular period of time.

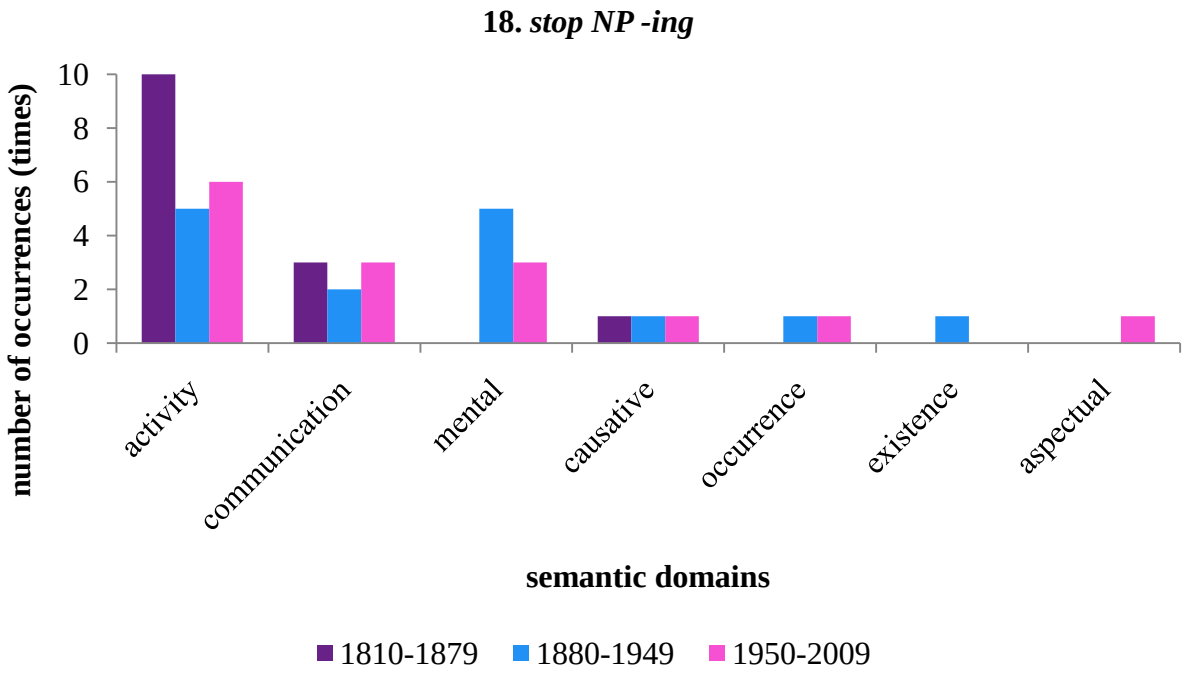
In 1880-1949, the activity verbs domain is still dominating especially for *stop NP from -ing* with a high number of occurrences. Some examples are *getting, doing, trying, selling, throwing* and *marrying*. For s*top NP from -ing*, there are not much varieties in the semantic domains for the *-ing* forms as most of them have low number of *-ing* forms being classified under them or are completely absent (causative and aspectual verbs domains). This trend differs for *stop NP -ing*. There are a high number of verbs being classified under mental verbs domain like *caring, looking, feeling, thinking* and *annoying.* But once again, there is no occurrence for aspectual verbs domain which is similar to *stop NP -ing*.

In 1950-2009, the dominating semantic domain is still activity verbs with a high number of *-ing* forms occurrences especially for *stop NP from -ing*. Examples of activity verbs in this time period is *getting, taking, dipping, passing, losing, pedalling, reaching, buying, winning killing* and *brushing*. This activity verbs domain has less number of occurrences for *stop NP -ing* though in comparison to the other domains, it is still dominating. For both *stop NP from -ing* and *stop NP -ing* variants, the number of *-ing* forms in the communication verbs domain (*telling, singing* and *saying*) and causative verbs domain (*exploding* and *destroying*) is low based on the number of occurrence times. For existence verbs domain, no verbs occur in this distinctive domain.

Based on the summary from Figure 18 (Graphs 17 and 18) below, the *-ing* forms for the *stop* complementation clauses have similar semantic preference with the *prevent* complementation clauses which is activity verbs. For 1810-1879, the most attracted collexemes lie within the activity verbs domain. This domain remains the same in the next time period. In 1880-1949, the top 15 distinctive *-ing* forms have a much more diverse spread into the other semantic domains though activity verbs domain remains as the dominance especially for *stop NP from -ing*. In 1950-2009, the activity verbs domain maintains as the dominating domain. We can observe this phenomenon through *stop NP from -ing* where the number of verbs occurrences increase from 1810 to 2009. Nonetheless, these semantic domains reflect the development of the *stop* complementation clauses over three different time periods. This development can only be observed through the collostructional values of the diachronic distinctive collexeme analysis and not the raw frequencies. This is because the most frequent occurring verbs might have less chance to be attracted to the constructions as the distinctive collexeme analysis depends highly on probability.

Figure 18: Top 15 distinctive *-ing* forms in *stop NP from -ing* and *stop NP -ing* across three time periods: frequencies in times



## 6.8 Summary

This chapter accounted for the analysis of the distinctive collexemes diachronically based on the *-ing* forms collected from the *prevent* and *stop* complementation clauses in COHA. The diachronic distinctive collexeme analysis is an application where the verbs were tested to find out their attraction or repulsion to each construction by using the Fisher-Yates exact test. All the results from this analysis were discussed in details in Section 6.7 above.

To recap, both *prevent* and *stop* complementation clauses showed similarities in their semantic domain preferences based on the results of the top 15 most frequent *-ing* forms. In all three time periods (1810-1879, 1880-1949 and 1950-2009), the activity verbs domain was dominating in the top 15 most frequently occurring *-ing* forms in both *prevent* and *stop* complementation clauses. Other verb categories had low occurrences or were completely absent. However, these raw frequencies cannot be used to observe how both complementation clauses have developed across time because they do not show how strong the collexemes are attracted or repelled to the with *from* and without *from* variants. The collexemes strengths were demonstrated through their collostructional values which can only be obtained through the diachronic distinctive collexeme analysis.

Nevertheless, the results from the diachronic distinctive collexemes analysis did not show much change in the semantic domains for both *prevent* and *stop* complementation clauses. Both verbs had similar trends once again. In 1810-1879, activity verbs domain was the dominating category. This category remained as dominant in 1880-1949 but the numbers of distinctive collexemes

have decreased. In 1950-2009, the activity verbs domain was being favoured as the number of distinctive collexemes have increased back.

In summary, both *prevent* and *stop* complementation clauses favoured activity verbs strongly whether in their most frequent list or distinctive list. This can be observed through the high number of occurrences in their *-ing* forms when used with *from* or without *from* and especially towards the 21st century. This is an expected result as activity verbs are the most common verbs in the English language. According to Biber et al. (1999: 365), English speakers make a lot of distinctions among activity verbs in comparison to other domains whether in writing or conversation. They commonly talk about a wide range of physical activities by using activity verbs. Thus, this is the reason as why activity verbs domain was dominating in both types of complementation clauses.

This diachronic distinctive collexeme analysis did not reveal much about the development of the complementation clauses – it has only demonstrated that activity verbs domain was the favourite domain in both the frequent and distinctive lists. There were no changes in the favourite domain over time. Certain collostructional values in some of the time periods from the analysis were too small due to low number of verbs occurrences. This has caused many difficulties during the interpretation of those results. Overall, the results from this diachronic distinctive collexeme analysis were not very meaningful to this study. The methodology to analyse the results has to be altered for improvement.

The following chapter summarises the results from all the conducted analyses and discusses the research significance of this study. The methodological limitations which were faced throughout the study are presented and I will give some suggestions for future directions.

# CHAPTER 7

# CONCLUSION

## 7.1 Overview

This chapter summarises my entire research project, draws some conclusions and highlights limitations and areas for future research. In Section 7.2, a brief summary of all the results obtained in the analyses from Chapters 5 and 6 is draw together. There were a total of four different analyses conducted in this study: (1) overall frequency per decade, (2) proportional values, (3) length of noun phrases, and (4) diachronic distinctive collexeme analysis. The following section (7.3) assembles the research significance of this study. Some methodological issues which were faced during the study will be examined in Section 7.4. The chapter ends with some suggestions for future research (Section 7.5).

## 7.2 Summary of principal findings

In this study of the *prevent* and *stop* complementation clauses in 19$^{th}$, 20$^{th}$ and 21$^{st}$ century American English based on COHA, there are a total of four different analyses which were conducted and they are as follows:

(i) Overall frequency per decade for the *prevent* and *stop* complementation clauses when used with and without *from* (Section 5.2.1, Chapter 5)

(ii) Comparison of the proportional values between the with *from* variant and the without *from* variant for *prevent* and *stop* (Section 5.2.2, Chapter 5)

(iii) Measurement of the average number of words within the length of noun phrases (Section 5.3, Chapter 5)

(iv) Diachronic distinctive collexeme analysis based on the *-ing* forms of the *prevent* and *stop* complementation clauses (Section 6.7, Chapter 6)

In the first analysis, the raw frequencies of the *prevent* and *stop* complementation clauses (with and without *from* variants) which were obtained from COHA were converted to per million words for standardisation as each decade has different number of samples. Then, four individual graphs were plotted (*prevent / stop + from* and *prevent / stop – from*) and interpreted based on the diachronic development of the complementation clauses.

In the second analysis, the raw frequencies of both complementation clauses were calculated in proportions and converted to percentage. The proportional values of the with *from* variant and the without *from* variant for *prevent* were demonstrated in one graph for comparison purpose. This step was repeated for *stop* and the results were interpreted based on the different stages of increasing and decreasing trends.

In the third analysis, the numbers of words within each noun phrase in every decade for both *prevent* and *stop* complementation clauses were calculated. The average numbers of words for different variants of the complementation clauses were used to plot four different graphs in order

to test Rohdenburg's Complexity Principle (1996) which is the favouring of the preposition or complementiser *from* in complex environments.

In the final analysis, the *-ing* forms of the complementation clauses were used in the diachronic distinctive collexeme analysis to test the attraction or repulsion of the *-ing* forms to the constructions. With the obtained collostructional values, the top 15 most distinctive *-ing* forms were classified into different semantic domains for the observation of their semantic domain preferences.

The basic assumption at the beginning of the study was shifts over time in frequency of the *prevent* and *stop* complementation clauses from COHA indicate linguistic change in progress. The hypothesis which I have made was the with *from* variant will dominate in American English for both verbs because American English marks negative orientation for verbs of negative causation by the use of *from* (Rohdenburg, 2009: 211). The *prevent* complementation clauses will experience a more stable trend while the *stop* complementation clauses will experience slower development. The results from the analyses have showed evidence which supported the hypothesis that in general especially towards the 21$^{st}$ century, the with *from* variant has become a dominating trend for both verbs. The different increasing and decreasing trends based on the overall frequency distribution of the *prevent* and *stop* complementation clauses indicated that American English has changed over the past 200 years. I conclude from my observation on the diachronic results that the process of language change has occurred in American English from

the 19th century to the 21st century. This process is expected to take place continuously in the future.

## 7.3 Research significance of this study

Although this study is a small research project on the development of the *prevent* and *stop* complementation clauses in 19th, 20th and 21st century American English, it still contributes in some ways to the present-day research of English language. Below are its contributions from my perspective.

This study was a detailed investigation of the non-finite complementation clauses focusing on two semantically comparable verbs (*prevent* and *stop*) in American English from the 19th century to the beginning of the 21st century based on COHA. As there were no previous studies which focus on the diachronic development of the *prevent* and *stop* complementation clauses in American English across the 19th, 20th and 21st centuries, this study has filled in the gap by using the empirical corpus-based methodology. The study has sought evidence to support Mair's prediction (2002) that American English will maintain the prepositional variant as dominant. The analyses that were conducted have showed quantitative results based on the distribution of frequencies. From the beginning of the 19th century to the 21st century, American English has favoured only one variant which is the with *from* variant especially for the *prevent* complementation clauses. This trend was less clear with the *stop* complementation clauses. Nevertheless, towards the 21st century, the *stop* complementation clauses have also chose to favour the with *from* variant only.

This study had its own distinctive analyses on different grammatical features within the *prevent* and *stop* complementation clauses especially like the diachronic distinctive collexeme analysis which was inspired by Hilpert (2006). No previous studies, to my knowledge, have conducted the diachronic distinctive collexeme analysis by using the *-ing* forms from both *prevent* and *stop* complementation clauses and classified them into different semantic domains based on Biber et al.'s classification of verbs (1999). This analysis was to explore any changes of their semantic domain preferences. It was considered unique although the results did not show any changes for the preferences of the activity verb domain. It was expected as activity verbs are the most commonly used verbs in the English language. The comparison analysis of the proportional values of the with and without *from* variants was much useful to this study. It has provided evidence to support my hypothesis. The with *from* variant for both verbs has become a dominating trend. *Prevent NP from -ing* showed a decreasing trend towards the 20th century but increased back towards the 21st century. Since then, it has almost taken over *prevent NP -ing* completely. *Stop NP from -ing* showed fluctuation in the 19th century until the 20th century. However, towards the 21st century, its trend started increasing and became a more common trend. These results could now contribute to structure language teaching materials or to revise existing grammar books, especially those aimed at non-native English speakers.

In short, even though some of the analysis results such as the length of noun phrases did not show strong evidence to support my hypothesis, they have still lead new directions to future research for any other non-finite complementation clauses or similar syntactic patterns. This study has also enabled us to gain deeper understanding towards the process of language change in general through the analysis of non-finite complementation clauses.

## 7.4 Methodological limitations

As expected, every research will face limitations in its methodology. After conducting all analyses for this study, I draw together methodological limitations which I have faced that might have caused some differences in the results.

The main challenge faced is the limitation of the recall of search queries (refer Section 4.3, Chapter 4). The queries were set to a maximum of three nouns only. The maximum use of other grammatical features such as adjectives was two only. They also did not include embedded clauses. This study was limited to two variants of the non-finite complementation clauses (with and without *from*) and focused only on two semantically comparable verbs, *prevent* and *stop* (refer Chapter 1). Other verbs which were categorised under the same semantic domain or might have shown similar results to support the development of *prevent* and *stop*, were not included.

There were not enough sample constructions for the beginning of the 19$^{th}$ century (1810-1860) based on the raw frequencies which were obtained from COHA (refer Section 5.2.1, Chapter 5). Thus, the development of the *prevent* and *stop* complementation clauses could not be observed clearly for the 19$^{th}$ century. In addition, the study focused on written genre only because COHA is a corpus of four different written genres (refer Section 3.5, Chapter 3). This issue on genre has limited further investigation on spoken genre of which might have uncovered some new results to the process of language change.

In the diachronic distinctive collexeme analysis, not all the *-ing* forms were classified into their respective semantic domains due to time constraint – only the top 15 *-ing* forms in the results were chosen (refer Section 6.7, Chapter 6). During the process of classification of verbs, there were some *-ing* forms that did not belong to any of the semantic domain – this phenomenon has caused much confusion. The process of classification of verbs had to rely only on the meaning of the context as there were no clear-cuts to it.

## 7.5 Suggestions for future research

The methodological limitations discussed in Section 7.4 can be improved. Hence, I would like to suggest some directions for future research.

First, the search queries should not limit the amount of nouns or other grammatical features during formation in order to capture longer noun phrases and embedded clauses. Deeper investigation can be conducted with more varieties of data to test Rohdenburg's Complexity Principle (1996). This will create an easier link to look upon the other linguistic patterns of linguistic change in progress such as colloquialization and grammaticalization. Extension to the current research can be conducted, probably including other complementation clauses such as the 'with and without *to*' variants for more diverse comparison. This includes other semantically comparable verbs such as *help to, start to* and *begin to*.

In order to conduct a more thorough investigation based on the diachronic American English data, the Corpus of Contemporary American English (COCA) and Google Books (American English) Corpus can be used. COCA has approximately 400+ million words ranging from 1990-2010 while Google Books has approximately 155 billion words ranging from 1810-2009. Both corpora use same tagset as COHA, thus the same search queries can be used. A new study focusing on the spoken genre can be conducted in order to compare the similarities and differences against the written genre for each distinctive grammatical feature within the complementation clauses.

Last but not least, more time should be spent on the diachronic distinctive collexeme analysis for the classification of the *-ing* forms into different semantic domains. More *-ing* forms should be classified in order to obtain more data which may feature new and interesting results.

To conclude, this study has presented some exciting results based on COHA which were not previously noted in the literature. In general, the 19$^{th}$, 20$^{th}$ and 21$^{st}$ century American English favours *prevent* and *stop NP from -ing* over *prevent* and *stop NP -ing*. Hence, this study has achieved its ultimate goal to uncover interesting frequency changes of the development of the *prevent* and *stop* complementation clauses in American English over a period of 200 years. It has also showed that in American English, linguistic change is still in progress from the 19$^{th}$ century until present day due to continuous process of language change.

# REFERENCES


Aarts, B. (1990). *Prevent*-type Verbs in a GB Framework. *UCL Working Papers in Linguistics, 2*, 147-164.

Aiello, L. C. (2010). Working Memory and the Evolution of Modern Thinking. Wenner-Gren Symposium Supplement 1. *Current Anthropology, 51*(1), 3-4.

Aitchison, J. (1991). *Language Change: Progress or Decay?* (2nd edition). Cambridge: Cambridge University Press.

Algeo, J. (1988). British and American Grammatical Differences. *International Journal of Lexicography, 1*, 1-31.

Algeo, J. (2006). *British or American English? A Handbook of Word and Grammar Patterns.* Cambrige: Cambridge University Press.

Babováková, P. (2005). *The Complements of Prevent.* Unpublished Master thesis, University of Tampere, Tampere.

Barber, C. (1964). *Linguistic Change in Present-Day English.* London and Edinburgh: Oliver and Boyd.

Barlow, M. (1998). *A Corpus of Spoken Professional American English.* Houston, Texas: Athelstan.

Bauer, L. (1994). *Watching English Change: An Introduction to the Study of Linguistic Change in Standard Englishes in the Twentieth Century.* London: Longman.

Berglund, Y. (1997). Future in Present-Day English: Corpus-Based Evidence on the Rivalry of Expressions. *ICAME Journal, 21*, 7-19.

Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing, 8*(4), 243-257.

Biber, D. (2003). Compressed Noun-Phrase Structures in Newspaper Discourse: The Competing Demands of Popularization vs. Economy. In J. Aitchison and D. M. Lewis (Eds.) *New Media Language* (pp. 169-181). London: Routledge.

Biber, D. and Clark, V. (2002). Historical Shifts in Modification Patterns with Complex Noun Phrase Structures. In T. Fanego, M. López-Couso and J. Pérez-Guerra (Eds.) *English Historical Morphology. Selected Papers from II ICEHL, Santiago de Compostela, 7-11 September 2000* (pp. 43-66). Amsterdam: Benjamins.

Biber, D., Conrad, S. and Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use.* Cambridge: Cambridge University Press.

Biber, D. and Finegan, E. (1997/2001). Diachronic Relations among Speech-Based and Written Registers in English. In T. Nevalainen and L. Kahlas-Tarkka (Eds.) *To Explain the Present: Studies in the Changing English Language in Honour of Matti Rissanen* (pp. 253-275). Helsinki: Société Néo-philologique. Reprinted in S. Conrad and D. Biber (Eds.) *Variation in English: Multidimensional Studies* (pp. 66-83). London: Longman.

Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Longman Grammar of Spoken and Written English.* London: Longman.

Binnick, R. I. (1971). Will and Be Going To. *Chicago Linguistic Society, 7*, 40-51.

Bloomfield, L. (1933). *Language.* New York: Holt, Rinehart and Winston.

Bolinger, D. (1968). Entailment and the Meaning of Structures. *Glossa, 2*, 119-127.

Bonfante, G. (1946). 'Indo-Hittite' and Areal Languages. *American Journal of Philology, 67*, 289-310.

Brinton, L. J. and Traugott, E. C. (2005). *Lexicalization and Language Change.* Cambridge: Cambridge University Press.

Brown, L. (Ed.) (1993). *The New Shorter Oxford English Dictionary on Historical Principles* (Volumes 1-2). Oxford: Oxford University Press.

Bybee, J., Perkins, R. and Pagliuca, W. (1994). *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World.* Chicago: University of Chicago Press.

Carnie, A. (2007). *Syntax* (2nd edition). Oxford: Blackwell.

Chomsky, N. A. (1957). *Syntactic Structures.* The Hague: Mounton.

Chomsky, N. A. (1962). A Transformation Approach to Syntax. Paper given and transcript of discussion in A. A. Hill (Ed.) *Third Texas Conference on Problems of Linguistic Analysis in English, 9-12 May 1958* (pp. 124-169). Austin, Texas: University of Texas.

Church, K. W., Gale, W., Hanks, P. and Hindle, D. (1991). Using Statistics in Lexical Analysis. In U. Zernik (Ed.) *Lexical Acquisition: Exploiting On-line Resources to Build Up a Lexicon* (pp. 115-164). Hillsdale, NJ: Lawrence Erlbaum.

Close, R. A. (1977) Some Observations on the Meaning and Function of Verb Phrases Having Future Time Reference. In W. D. Bald and R. Ilson (Eds.) *Studies in English Usage: The Resources of a Present-Day English Corpus for Linguistic Analysis* (pp. 125-56). Frankurt am Main: Peter Lang.

Comrie, B. (1989). *Language Universals and Linguistic Typology: Syntax and Morphology* (2nd edition). Oxford: Blackwell.

Croft, W. (2000). *Explaining Language Change: An Evolutionary Approach.* Essex: Pearson Education.

Crowley, T. (1989). *The Politics of Discourse: The Standard Language Question in British Cultural Debates.* London: Macmillan.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990-2008+): Design, Architecture and Linguistic Insights. *International Journal of Corpus Linguistics ,14*(2), 159-190.

Denison, D. (1998). Syntax. In S. Romaine (Ed.) *The Cambridge History of the English Language* (Volume 4: 1776-1997) (pp. 92-329). Cambridge: Cambridge University Press.

Dixon, R. M. W. (1991). *A New Approach to English Grammar on Semantic Principles.* Oxford: Clarendon Press.

Dixon, R. M. W. (1995). Complement Clauses and Complementation Strategies. In F. R. Palmer (Ed.) *Grammar and Meaning: Essays in Honour of Sir John Lyons* (pp. 175-220). Cambridge: Cambridge University Press.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics: Quantitative, Qualitative and Mixed Methodologies.* Oxford: Oxford University Press.

Elsness, J. (1994). On the Progression of the Progressive in Early Modern English. *ICAME Journal, 18*, 5-25.

Fang, A. C. (1992). Building a Corpus of Computer Science English. In J. Aarts, P. de Hann and N. Oostdik (Eds.) *English Language Corpora: Design, Analysis and Exploitation* (pp. 73-78). Amsterdam: Rodopi.

Firth, J. R. (1957). A Synopsis of Linguistic Theory 1930-1955. In *Studies in Linguistic Analysis* (pp. 1-32). Oxford: Philological Society. Reprinted in F. R. Palmer (Ed.) *Selected Papers of J. R. Firth 1952-1959.* London: Longman.

Francis, G., Hunston, S. and Manning, E. (1996). *Collins COBUILD Grammar Patterns 1: Verbs.* London: Harper Collins.

Francis, G., Hunston, S. and Manning, E. (1998). *Collins COBUILD Grammar Patterns II: Nouns and Adjectives.* London: Harper Collins.

Francis, W. N. (1965). A Standard Corpus of Edited Present-Day American English. *College English, 26*(4), 267-273.

Fries, C. (1945). *Teaching and Learning English as a Foreign Language.* Ann Arbor: University of Michigan Press.

Goldberg, A. E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure.* USA: University of Chicago Press.

Goldberg, A. E. (1996). Construction Grammar. In F. K. Brown and J. E. Millar (Eds.) *Concise Encyclopedia of Syntactic Theories* (pp. 68-71). Oxford: Pergamon.

Graddol, D. (1997). *The Future of English? A Guide to Forecasting the Popularity of the English Language in the 21st Century.* London: British Council.

Gries. S. Th. (to appear). Corpus Linguistics: Quantitative Methods. In C. A. Chapelle (Ed.) *The Encyclopedia of Applied Linguistics.* Oxford: Blackwell.

Gries, S. Th. and Stefanowitsch, A. (2004a). Extending Collostructional Analysis: A Corpus-Based Perspective on 'Alternations'. *International Journal of Corpus Linguistics, 9*(1), 97-129.

Gries, S. Th. and Stefanowitsch, A. (2004b). Co-varying Collexemes in the Into-Causative. In M. Achard and S. Kemmer (Eds.) *Language, Culture and Mind* (pp. 225-236). Stanford, CA: University of Chicago Press.

Haegeman, L. M. V. (1989). Be Going To and Will: A Pragmatic Account. *Journal of Linguistics, 25*, 291-317.

Harris, Z. (1957). Co-occurrence and Transformation in Linguistic Structure. *Language, 33*(3), 283-340.

Hawkins, J. (1990). A Parsing Theory of Word Order Universals. *Linguistic Inquiry, 21*, 223-261.

Hawkins, J. (1992). Syntactic Weight versus Information Structure in Word Order Variation. In J. Jacobs (Ed.) *Informationsstruktur und Grammatik* (Linguistische Berichte Sonderheft 4) (pp. 186-219). Opladen: Westdeutscher Verlag.

Heine, B. (1993). *Auxiliaries: Cognitive Forces and Grammaticalization.* Oxford: Oxford University Press.

Heyvaert, L., Rogiers, H. and Vermeylen, N. (2005). Pronominal Determiners in Gerundive Nominalization: A Case Study. *English Studies, 86*, 71-88.

Hilpert, M. (2006). Distinctive Collexeme Analysis and Diachrony. *Corpus Linguistics and Linguistic Theory, 2*(2), 243-256.

Hilpert, M. (to appear). *Diachronic Collostructional Analysis.* (Teaching material)

Hilpert, M. and Gries, S. Th. (2009) Assessing Frequency Changes in Multistage Diachronic Corpora: Applications for Historical Corpus Linguistics and the Study of Language Acquisition. *Literary and Linguistic Computing, 24*(4), 385-401.

Hoffmann, S., Evert, S., Smith, N., Lee, D. and Prytz, Y.B. (2008). *Corpus Linguistics with BNCweb – A Practical Guide.* Frankfurt am Main: Peter Lang.

Hoey, M. (1991). *Pattern of Lexis in Text.* Oxford: Oxford University Press.

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language.* London: Routledge.

Hopper, P. J. and Traugott, E. C. (2003). *Grammaticalization* (2nd edition). Cambridge: Cambridge University Press.

Howell, D. C. (2010). *Statistical Methods for Pscyhology* (International edition). Belmont, CA: Wadsworth Cengage Learning.

Howitt, D. and Cramer, D. (2011). *Introduction to Statistics in Psychology* (5th edition). Essex: Pearson.

Huddleston, R. (1984). *Introduction to the Grammar of English.* Cambridge: Cambridge University Press.

Huddleston, R. and Pullum, G. K. (2002). *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Huddleston, R. and Pullum, G. K. (2005). *A Student's Introduction to English Grammar.* Cambridge: Cambridge University Press.

Hundt, M. (2004). Animacy, Agentivity and the Spread of the Progressive in Modern English. *English Language and Linguistics, 8*(1), 47-69.

Hunston, S. and Francis, G. (2000). *Pattern Grammar: A Corpus-Driven Approach to the Lexical Grammar of English.* Amasterdam: John Benjamins.

Janda, R. D. and Joseph, B. D. (2003). On Language, Change and Language Change – Or, Of History, Linguistics and Historical Linguistics. In B. D. Joseph and R. D. Janda (Eds.) *The Handbook of Historical Linguistics* (pp. 4-180). Oxford: Blackwell.

Jesperson, O. (1909-1949). *A Modern English Grammar on Historical Principles.* London: Allen and Unwin.

Johansson, S. (1980). *Plural Attributive Nouns in Present-Day English.* Lund: CWK Gleerup.

Kachru, B. B. (1992). Teaching World Englishes. In B. B. Kachru (Ed.) *The Other Tongue: English across Cultures* (2nd edition) (pp. 355-366). USA: University of Illinois Press.

Kidd, E., Lieven, E. V. and Tomasello, M. (2005) The Acquisition of Complement Clause Constructions: A Sentence Repetition Study. In *Proceedings of the 32nd Stanford Child Language Research Forum* (pp. 50-59). Stanford, California.

Kim, K. K., Byun, E., Lee, S. K., Gaillard, W. D., Xu, B. and Theodore, W. H. (2011). Verbal Working Memory of Korean-English Bilinguals: An fMRI Study. *Journal of Neurolinguistics, 24,* 1-13.

King, J. and Just, A. (1991). Individual Differences in Syntactic Processing: The Role of Working Memory. *Journal of Memory and Language, 30,* 580-602.

Kjellmer, G. (1994). *A Dictionary of English Collocations: Based on the Brown Corpus* (Volume 1, 2 and 3). Oxford: Clarendon Press.

Lakoff, G. (1987). *Women, Fire and Dangerous Things: What Categories Reveal about the Mind.* Chicago and London: University of Chicago Press.

Leech, G. (1991). The State of the Art in Corpus Linguistics. In K. Aijmer and B. Altenberg (eds.) *English Corpus Linguistics: Studies in Honour of Jan Svartvik* (pp. 8-29). London: Longman.

Leech, G. (1992). 100 Million Words of English: the British National Corpus. *Language Research, 28*(1), 1-13.

Leech, G. and Smith, N. (2005). Extending the Possibilities of Corpus-Based Research on English in the Twentieth Century: A Prequel of LOB and FLOB. *ICAME Journal, 29,* 83-98.

Leech, G., Mair, C., Hundt, M. and Smith, N. (2009). *Change in Contemporary English: A grammatical Study.* Cambridge: Cambridge University Press.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation.* Chicago and London: University of Chicago Press.

Lin, X. (1999). *Cihui Yuyi he Jisuan Yuyanxue. [Lexical Semantics and Computational Linguistics]*. Beijing: Lingua Publishing House.

*Longman Dictionary of Contemporary English.* (1978). London: Longman.

Matthews, P. H. (2007). *Oxford Concise Dictionary of Linguistics* (2nd edition). Oxford: Oxford University Press.

Mair, C. (1995). Changing Patterns of Complementation, and Concomitant Grammaticalization, of the Verb *help* in Present-day British English. In B. Aarts, and C. F. Meyer (Eds.) *The Verb in Contemporary English: Theory and Description* (pp. 258-272). Cambridge: Cambridge University Press.

Mair, C. (2002). Three Changing Patterns of Verb Complementation in Late Modern English: A Real-time Study Based on Matching Text Corpora. *English Language and Linguistics, 6*(1), 105-132.

Mair, C. (2006). *Twentieth-century English: History, Variation and Standardization.* Cambridge: Cambridge University Press.

McEnery, T., Xiao, R. and Tono, Y. (2006). *Corpus-Based Language Studies: An Advanced Resource Book.* London: Routledge.

Miller, G., Galanter, E. and Pribram, K. (1960). *Plans and the Structure of Behaviour.* New York: Holt, Rinehart and Winston.

Nicholson, M. (1957). *A Dictionary of American – English Usage: Based on Fowler's Modern English Usage.* New York: Oxford University Press.

Otlogetswe, T. (2004). The BNC Design as a Model for a Setswana Language Corpus. Paper presented at the 7th Annual CLUK Research Colloquium, University of Birmingham, 6-7 January 2004. Retrieved 10 April 2011, from http://cs.bham.ac.uk/~mgl/cluk/titles.html

*Oxford English Dictionary.* (2010). Retrieved 3 February 2011, from http://www.oed.com/

Paris, G. (1868). Review of Schleicher 1868. *Revue critique d'histoire et de littérature, 6*(2), 241-244.

Pedersen, T. (1996). Fishing for Exactness. In *Proceedings of the Conference of South-Central SAS Users Group* (pp. 188-200). Austin, Texas.

Postal, P. (1974). *On Raising.* Cambridge, MA: Massachusetts Institute of Technology Press.

Potter, S. (1975). *Changing English* (2nd edition). London: Deutsch.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

Quirk, R. and Stein, G. (1990). *English in Use.* London: Longman.

Radford, A. (1988). *Transformational Grammar: A First Course.* Cambridge: Cambridge University Press.

Rohdenburg, G. (1995). On the Replacement of Finite Complement Clauses by Infinitives in English. *English Studies, 76*, 367-388.

Rohdenburg, G. (1996). Cognitive Complexity and Increased Grammatical Explicitness in English. *Cognitive Linguistics, 7*, 149-182.

Rohdenburg, G. (2006). Processing Complexity and Competing Sentential Variants in Present-Day English. In W. Kürschner and R. Rapp (Eds.*) Linguistik International: Festschrift für Heinrich Weber* (pp. 51-67). Lengerich: Pabst Science Publishers.

Rohdenburg, G. (2009). Nominal Complements. In G. Rohdenburg and J. Shlüter (Eds.) *One Language, Two Grammars? Differences between British and American English* (pp. 194-211). Cambridge: Cambridge University Press.

Rosembaum, P. S. (1967). *The Grammar of English Predicate Complement Constructions.* Cambridge, MA: Massachusetts Institute of Technology Press.

Rosenbach, A. (2002). *Genitive Variation in English: Conceptual Factors in Synchronic and Diachronic Studies.* Berlin and New York: Mouton de Gruyter.

Rudanko, J. (2002). *Complements and Constructions: Corpus-based Studies on Sentential Complements in English in Recent Centuries.* Lanham, Maryland: University Press of America.

Rudanko, J. (2006). Watching English Grammar Change: A Case Study on Complement Selection in British and American English. *English Language and Linguistics, 10*(1), 31-48.

Santos, D. (1996). *Tense and Aspect in English and Portuguese: A Contrastive Semantical Study.* Unpublished PhD thesis, Universidad Tecnica de Lisboa, Lisbon.

Sellgren, E. (2007). *A Corpus-based Study of the Complements of Prevent in the 18th, 19th, and 20th Centuries.* Unpublished Master thesis, University of Tampere, Tampere.

Sellgren, E. (2009). *Tracing the Sentential Complements of Prevent through Centuries.* Retrieved from http://ucrel.lancs.ac.uk/publications/cl2009/214_FullPaper.docx.

Sellgren, E. (2010). *Prevent* and the battle of the *-ing* clauses. In U. Lenker, J. Huber, R. Mailhammer (Eds.) *English Historical Linguistics 2008 Volume 1: The History of English Verbal and Nominal Constructions* (pp. 45-62). Amsterdam: Benjamins.

Shimazumi, M. and Berber-Sardinha, A. (1996). Approaching the Assessment of Performance Unit (APU) Archive of Schoolchildren's Writing from the Point of View of Corpus Linguistics. Paper presented at the TALC 1996 Conference, Lancaster University, 11 August 1996.

Sinclair, J. M. (Ed.) (1987). *Collins Cobuild English Language Dictionary.* London and Glasgow: Collins.

Sinclair, J. M. (Ed.) (1990). *Collins Cobuild English Grammar.* London: Harper Collins.

Sinclair, J. M. (1991). *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

Sinclair, J. M. (1995). Paper presented at XI Encontro da Associaç ão Portuguesa de Linguística, Lisbon, 2-4 October 1995.

Sinclair, J. M. (1996). *EAGLES: Preliminary Recommendations on Corpus Typology*. Retrieved 10 January 2011, from http://www.ilc.cnr.it/EAGLES96/corpustyp/corpustyp.html

Stefanowitsch, A. (2005). *Collostructional Analysis Workshop (Version 1.0*). Retrieved 15 March 2011, from http://www-user.uni-bremen.de/~anatol/crs/caw/index.html

Stefanowitsch, A. (2006). Distinctive Collexeme and Diachrony: A Comment. *Corpus Linguistics and Linguistic Theory, 2*(2), 257-262.

Stefanowitsch, A. and Gries, S. Th. (2003). Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics, 8*(2), 209-243.

Strang, B. M. H. (1982). Some Aspects of the History of the Be + ing Construction. In J. Anderson (ed.) *Language Form and Linguistic Variation: Papers Dedicated to Angus Mcintosh* (pp. 427-474). Amsterdam: Benjamins.

Stubbs, M. (1996). *Text and Corpus Analysis.* Oxford: Blackwell.

Svartvik, J. and Leech, G. (2006). *English: One Tongue, Many Voices.* Basingstoke: Palgrave Macmillan.

*The American Heritage Dictionary of the English Language*. (3rd edition). (1992). Boston: Houghton Mifflin.

The English Language Institute University of Michigan (2002). *MICASE Manual: The Michigan Corpus of Academic Spoken English* (Version 3). Retrieved 10 April 2011, from http://micase.elicorpora.info/micase-manual-pdf

*The Daily Mail* and *The Mail on Sunday*. (1993-2000). (CD-ROM). Chadwyck-Healey.

Tognini-Bonelli, E. (2001). *Corpus Linguistics at Work.* Amsterdam: John Benjamins.

Van Ek, J. A. (1966). *Four Complementary Structures of Predication in Contemporary British English* (an inventory). Groningen: Wolters.

Van Oosten, J. H. (1984). *The Nature of Subjects, Topics and Agents: A Cognitive Explanation*. Bloomington: Indiana University Linguistics Club (IULC) Publications.

Visser, F. Th. (1973). *An Historical Syntax of the English Language (Part 3).* Leiden: Brill.

Vosberg, U. (2009). Non-Finite Complements. In G. Rohdenburg and J. Schlüter (Eds.) *One Language, Two Grammars? Differences between British and American English* (pp. 212-227). New York: Cambridge University Press.

Watson, O. (1976). *Longman Modern English Dictionary* (2nd edition). London: Longman.

*Webster's Dictionary of English Usage*. (1989). Springfield, MA: Merriam-Webster.

*Webster's New Collegiate Dictionary* (8th edition). (1973). London: G. Bell and Sons Ltd.

*Webster's New World Dictionary of American English* (3rd college edition). (1988). New York: Macmillan General Reference.

Wekker, H. C. (1976). *The Expression of Future Time in Contemporary British English.* Amsterdam: North Holland.

Yamamoto, M. (1999). *Animacy and Reference: A Cognitive Approach to Corpus Linguistics.* Amsterdam/Philadelphia: Benjamins.

# APPENDICES

## APPENDIX 1: Corpora sources

**ARCHER – A Representative Corpus of Historical English Registers**

| | |
|---|---|
| Developer | Douglas Biber (Northern Arizona University), Edward Finegan (University of Southern California) and 14 universities in 7 countries |
| Period | 1650 – 1999 |
| Size | 1.7 million words |
| Language | British and American English |
| Contents | Diachronic written language; 11 categories sampled from 8 historical periods starting from Early Modern English; more than 1037 texts |
| Annotation | Untagged and tagged version, morphological tagging |
| Website | http://www.llc.manchester.ac.uk/research/projects/archer/ |

**BoE – BANK OF ENGLISH (Collins Cobuild Corpus)**

| | |
|---|---|
| Developer | John Sinclair and his team (University of Birmingham and Harper-Collins) |
| Period | 1990 – present day |
| Size | 500 million words |
| Language | Standard English (70% British English, 20% American English, 10% other varieties) |
| Contents | 75% of written texts; 25% of spoken language; contains entire texts rather than samples |
| Annotation | POS tagging (BoE tagset) |
| Website | http://www.harpercollins.co.uk/about-harpercollins/Imprints/collins/Pages/Collins.aspx |

**BASELINE 1900 – Oxford English Dictionary Baseline 1900**

| | |
|---|---|
| Developer | - |
| Period | 1896 – 1905 |
| Size | 993 000 words |
| Language | British English |
| Contents | Quotations from Oxford English Dictionary (1989, 2nd edition) |
| Annotation | - |
| Website | Mair, C. (2006). *Twentieth-century English: History, Variation and Standardization* (pp. 210-212). Cambridge: Cambridge University Press. |

**BNC – The British National Corpus**

| | |
|---|---|
| Developer | An academic-industrial consortium from Oxford University Press (Addison-Wesley Longman and Larousse Kingfisher Chambers, Oxford University Computing Services, University Centre for Computer Corpus Research on Language at Lancaster University, and British Library Research and Development Department) |
| Period | 1980s – 1993 |
| Size | 100 million words |
| Language | Contemporary British English |
| Contents | 90% written and 10% spoken language; more than 4000 texts |
| Annotation | Textual markup, discourse annotation, POS tagging (CLAWS tagset) |
| Website | http://www.natcorp.ox.ac.uk/ |

**BROWN – Standard Corpus of Present-Day Edited American English for Use with Digital Computers**

| | |
|---|---|
| Developer | Nelson Francis and Henry Kučera (Brown University) |
| Period | 1961 |
| Size | 1 million words |
| Language | Contemporary American English |
| Contents | Written language; 15 text categories; 500 text samples of 2000 words each |
| Annotation | POS tagging (CLAWS tagset) |
| Website | http://cqpweb.lancs.ac.uk/ |

**CLMET – Corpus of Late Modern English Texts**

| | |
|---|---|
| Developer | Henrik de Smet (University of Leuven) |
| Period | 1710 – 1920 |
| Size | 10 million words |
| Language | British English |
| Contents | Diachronic written language; a collection of texts drawn from the Project Gutenberg and the Oxford Text Archive; each author contributes to no more than 200 000 words of text |
| Annotation | POS tagging |
| Website | https://perswww.kuleuven.be/~u0044428/ |

**COCA – Corpus of Contemporary American English**

| | |
|---|---|
| Developer | Mark Davies (Brigham Young University) |
| Period | 1990 – present day |
| Size | 410+ million words |
| Language | Contemporary American English |
| Contents | 20 million words each year divided equally between spoken language and 4 text categories; more than 200 000 texts |
| Annotation | POS tagging (CLAWS tagset) |
| Website | http://corpus.byu.edu/coca/ |

**COHA – Corpus of Historical American English**

| | |
|---|---|
| Developer | Mark Davies (Brigham Young University) |
| Period | 1810 – 2009 |
| Size | 400+ million words |
| Language | American English |
| Contents | Diachronic written language; the number of words in each year is divided equally between 4 text categories |
| Annotation | POS tagging (CLAWS tagset) |
| Website | http://corpus.byu.edu/coha/ |

**CSPA – Corpus of Professional Spoken American English**

| | |
|---|---|
| Developer | Michael Barlow (Athelstan and Rice University) |
| Period | 1994 – 1998 |
| Size | 2 million words (2 sub-corpora of 1 million words each) |
| Language | Contemporary American English |
| Contents | Academic discourse and White House briefings; short interchanges of 400 speakers |
| Annotation | Untagged and tagged version (POS tagging – CLAWS tagset) |
| Website | http://www.athel.com/cpsa.html |

**FLOB – Freiburg-LOB Corpus of British English**

| | |
|---|---|
| Developer | Christian Mair (University of Freiburg) |
| Period | 1991 – 1992 |
| Size | 1 million words |
| Language | Contemporary British English |
| Contents | Written language; 15 text categories; 500 text samples of 2000 words each |
| Annotation | Untagged and tagged version (POS tagging – CLAWS tagset) |
| Website | http://cqpweb.lancs.ac.uk/ |

**FROWN – Freiburg-Brown Corpus of American English**

| | |
|---|---|
| Developer | Christian Mair (University of Freiburg) |
| Period | 1991 – 1992 |
| Size | 1 million words |
| Language | Contemporary American English |
| Contents | Written language; 15 text categories; 500 text samples of 2000 words each |
| Annotation | Untagged and tagged version (POS tagging – CLAWS tagset) |
| Website | http://cqpweb.lancs.ac.uk/ |

**GOOGLE BOOKS – Google Books (American English) Corpus**

| | |
|---|---|
| Developer | Mark Davies (Brigham Young University) |
| Period | 1810 – 2009 |
| Size | 155 billion words |
| Language | American English |
| Contents | Diachronic written language; more than 1.3 million books |
| Annotation | POS tagging (CLAWS tagset) |
| Website | http://googlebooks.byu.edu/ |

**HELSINKI – Helsinki Corpus of English Texts: Diachronic Part**

| | |
|---|---|
| Developer | M. Rissanen, O. Ihalainen and M. Kytö (University of Helsinki) |
| Period | c. 750 – c. 1700 |
| Size | 1.5 million words |
| Language | British English |
| Contents | Diachronic written language; Old, Middle and Early Modern English texts sampled from 11 periods ($8^{th}$ – $18^{th}$ centuries); 400 text samples; additional of text samples from Old Scots and Old American English |
| Annotation | Textual markup (ASCII codes) |
| Website | http://khnt.hit.uib.no/icame/manuals/hc/index.htm |

**HKUST – Hong Kong University of Science and Technology Computer Science Corpus**

| | |
|---|---|
| Developer | Alex Chengyu Fang (University College London) |
| Period | From 1985 |
| Size | 1 million words |
| Language | American English |
| Contents | Written text samples from the Computer Science reading lists for first-year students in 7 tertiary institutions in Hong Kong |
| Annotation | Text categorisation, tagged version (AUTASYS tagset) |
| Website | Fang, A. C. (1992). Building a Corpus of Computer Science English. In J. Aarts, P. de Hann and N. Oostdik (Eds.) *English Language Corpora: Design, Analysis and Exploitation* (pp. 73-78). Amsterdam: Rodopi. |

**ICE-GB – British Component of the International Corpus of English**

| | |
|---|---|
| Developer | Co-ordinated by Gerald Nelson (Chinese University of Hong Kong) |
| Period | From 1990s |
| Size | 1 million words |
| Language | Contemporary British English |
| Contents | Written and spoken language; 200 written and 300 spoken texts from 32 categories |
| Annotation | Textual markup, discourse annotation, POS tagging, syntactic parsing, hand checked |
| Website | http://www.ucl.ac.uk/english-usage/projects/ice-gb/ |

**LLC London – London-Lund Corpus of Spoken English**

| | |
|---|---|
| Developer | Randolph Quirk and Sidney Greenbaum (University College London) and Jan Svartvik (Lund University) |
| Period | 1953 – 1987 |
| Size | 500 000 words |
| Language | British English |
| Contents | Derived from 2 projects – Survey of English Usage (SEU, 1959, University College London) and Survey of Spoken English (SSE, 1975, Lund University); 100 spoken texts of 5000 words each |
| Annotation | Prosodic and discourse annotation |
| Website | http://khnt.hit.uib.no/icame/manuals/londlund/index.htm |

**LOB – Lancaster-Oslo-Bergen Corpus**

| | |
|---|---|
| Developer | Geoffrey Leech (Lancaster University), Stig Johansson (University of Oslo) and Knut Hofland (University of Bergen) |
| Period | 1961 |
| Size | 1 million words |
| Language | Contemporary British English |
| Contents | Written language; 15 text categories; 500 text samples of 2000 words each |
| Annotation | Untagged and tagged version (POS tagging – CLAWS tagset) |
| Website | http://cqpweb.lancs.ac.uk/ |

**LSWE – Longman Spoken and Written English Corpus**

| | |
|---|---|
| Developer | Douglas Biber, Susan Conrad, Marie Helt and Erika Konrad (Northern Arizona University), |
| Period | 1980s onwards |
| Size | 40 million words |
| Language | Contemporary British and American English |
| Contents | Written and spoken texts from 4 categories; 37000 text samples |
| Annotation | POS tagging |
| Website | Biber, D., Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999). *Longman Grammar of Spoken and Written English* (pp. 24-28). London: Longman. |

**MICASE – Michigan Corpus of Academic Spoken English**

| | |
|---|---|
| Developer | R. C. Simpson, S. L. Briggs, J. Ovens and J. M. Swales (University of Michigan) |
| Period | 1997 – present day |
| Size | 1.7 million words |
| Language | Contemporary American English |
| Contents | Transcripts and audio files of academic speech within the University of Michigan |
| Annotation | Discourse annotation |
| Website | http://micase.elicorpora.info/micase-manual-pdf |

**PPCEME – Penn-Helsinki Parsed Corpus of Early Modern English**

| | |
|---|---|
| Developer | Anthony Kroch and Beatrice Santorini (University of Pennsylvania) |
| Period | 1500 – 1710 |
| Size | 1.8 million words |
| Language | Early Modern English |
| Contents | 229 prose text samples |
| Annotation | Syntactic annotation (parsing), POS tagging |
| Website | http://www.ling.upenn.edu/hist-corpora/ |

**TIME – TIME Corpus**

| | |
|---|---|
| Developer | Mark Davies (Brigham Young University) |
| Period | 1923 – present day |
| Size | 100 million words |
| Language | Contemporary American English |
| Contents | A collection of TIME magazines from 1923 to present day |
| Annotation | POS tagging (CLAWS tagset) |
| Website | http://corpus.byu.edu/time/ |

## APPENDIX 2: Abbreviation of recorded usage dates of English Language

| Acronym | Description | Period |
|---|---|---|
| OE | Old English | -1149 |
| LOE | Late Old English | 1000-1149 |
| ME | Middle English | 1150-1349 or, in some contexts, 1469 |
| LME | Late Middle English | 1350-1469 |
| L15 | Late fifteenth century | 1470-1499 |
| E16 | Early sixteenth century | 1500-1529 |
| M16 | Mid sixteenth century | 1530-1569 |
| L16 | Late sixteenth century | 1570-1599 |
| E17 | Early seventeenth century | 1600-1629 |
| M17 | Mid seventeenth century | 1630-1669 |
| L17 | Late seventeenth century | 1670-1699 |
| E18 | Early eighteenth century | 1700-1729 |
| M18 | Mid eighteenth century | 1730-1769 |
| L18 | Late eighteenth century | 1770-1729 |
| E19 | Early nineteenth century | 1800-1829 |
| M19 | Mid nineteenth century | 1830-1869 |
| L19 | Late nineteenth century | 1870-1899 |
| E20 | Early twentieth century | 1900-1929 |
| M20 | Mid twentieth century | 1930-1969 |
| L20 | Late twentieth century | 1970-1999 |
| E21 | Early twenty-first century | 2000-present time |

## APPENDIX 3: Additional tables

Table 24: Proportional values of *prevent NP from -ing* and *prevent NP -ing*: raw frequencies in instances and proportions in percentage (%) (cf. Figure 9, Graph 5, Section 5.2.2, Chapter 5)

| **Year** | **Total number of raw frequency (instances)** | ***prevent NP from -ing*** | | ***prevent NP -ing*** | |
|---|---|---|---|---|---|
| | | **Raw frequency (instances)** | **Proportions (%)** | **Raw frequency (instances)** | **Proportions (%)** |
| 1810 | 19 | 18 | 94.7 | 1 | 5.3 |
| 1820 | 273 | 246 | 90.1 | 27 | 9.9 |
| 1830 | 454 | 407 | 89.6 | 47 | 10.4 |
| 1840 | 421 | 367 | 87.2 | 54 | 12.8 |
| 1850 | 469 | 399 | 85.1 | 70 | 14.9 |
| 1860 | 469 | 394 | 84.0 | 75 | 16.0 |
| 1870 | 545 | 452 | 82.9 | 93 | 17.1 |
| 1880 | 560 | 479 | 85.5 | 81 | 14.5 |
| 1890 | 556 | 472 | 84.9 | 84 | 15.1 |
| 1900 | 540 | 433 | 80.2 | 107 | 19.8 |
| 1910 | 551 | 462 | 83.8 | 89 | 16.2 |
| 1920 | 635 | 547 | 86.1 | 88 | 13.9 |
| 1930 | 529 | 474 | 89.6 | 55 | 10.4 |
| 1940 | 532 | 494 | 92.9 | 38 | 7.1 |
| 1950 | 490 | 450 | 91.8 | 40 | 8.2 |
| 1960 | 488 | 446 | 91.4 | 42 | 8.6 |
| 1970 | 512 | 470 | 91.8 | 42 | 8.2 |
| 1980 | 528 | 506 | 95.8 | 22 | 4.2 |
| 1990 | 594 | 572 | 96.3 | 22 | 3.7 |
| 2000 | 627 | 609 | 97.1 | 18 | 2.9 |

Table 25: Proportional values of *stop NP from -ing* and *stop NP -ing*: raw frequencies in instances and proportions in percentage (%) (cf. Figure 9, Graph 6, Section 5.2.2, Chapter 5)

| **Year** | **Total number of raw frequency (instances)** | ***stop NP from -ing*** | | ***stop NP -ing*** | |
|---|---|---|---|---|---|
| | | **Raw frequency (instances)** | **Proportions (%)** | **Raw frequency (instances)** | **Proportions (%)** |
| 1810 | 0 | 0 | 0 | 0 | 0 |
| 1820 | 3 | 2 | 66.7 | 1 | 33.3 |
| 1830 | 6 | 5 | 83.3 | 1 | 16.7 |
| 1840 | 6 | 4 | 66.7 | 2 | 33.3 |
| 1850 | 9 | 5 | 55.6 | 4 | 44.4 |
| 1860 | 8 | 4 | 50.0 | 4 | 50.0 |
| 1870 | 18 | 12 | 66.7 | 6 | 33.3 |
| 1880 | 22 | 10 | 45.5 | 12 | 54.5 |
| 1890 | 30 | 15 | 50.0 | 15 | 50.0 |
| 1900 | 37 | 25 | 67.6 | 12 | 32.4 |
| 1910 | 57 | 31 | 54.4 | 26 | 45.6 |
| 1920 | 82 | 58 | 70.7 | 24 | 29.3 |
| 1930 | 90 | 60 | 66.7 | 30 | 33.3 |
| 1940 | 90 | 69 | 76.7 | 21 | 23.3 |
| 1950 | 112 | 83 | 74.1 | 29 | 25.9 |
| 1960 | 168 | 131 | 78.0 | 37 | 22.0 |
| 1970 | 192 | 143 | 74.5 | 49 | 25.5 |
| 1980 | 208 | 171 | 82.2 | 37 | 17.8 |
| 1990 | 311 | 278 | 89.4 | 33 | 10.6 |
| 2000 | 389 | 354 | 91.0 | 35 | 9.0 |

Table 26: Average number of words within the noun phrases for *prevent NP from -ing* and *prevent NP -ing*: frequencies in words (cf. Figure 10, Graphs 7 and 8, Section 5.3, Chapter 5)

| **Year** | ***prevent NP from -ing*** | | | ***prevent NP -ing*** | | |
|---|---|---|---|---|---|---|
| | **Raw frequency (instances)** | **Number of words within noun phrases** | **Average number of words (per words)** | **Raw frequency (instances)** | **Number of words within noun phrases** | **Average number of words (per words)** |
| 1810 | 18 | 20 | 1.11 | 1 | 2 | 1.00 |
| 1820 | 246 | 324 | 1.32 | 27 | 40 | 1.48 |
| 1830 | 407 | 581 | 1.43 | 47 | 68 | 1.45 |
| 1840 | 367 | 558 | 1.52 | 54 | 82 | 1.52 |
| 1850 | 399 | 583 | 1.46 | 70 | 100 | 1.43 |
| 1860 | 394 | 559 | 1.42 | 75 | 120 | 1.60 |
| 1870 | 452 | 661 | 1.46 | 93 | 123 | 1.32 |
| 1880 | 479 | 704 | 1.47 | 81 | 126 | 1.56 |
| 1890 | 472 | 701 | 1.49 | 84 | 116 | 1.38 |
| 1900 | 433 | 671 | 1.55 | 107 | 164 | 1.53 |
| 1910 | 462 | 733 | 1.59 | 89 | 141 | 1.58 |
| 1920 | 547 | 837 | 1.53 | 88 | 138 | 1.57 |
| 1930 | 474 | 740 | 1.56 | 55 | 92 | 1.67 |
| 1940 | 494 | 812 | 1.64 | 38 | 59 | 1.55 |
| 1950 | 450 | 747 | 1.66 | 40 | 58 | 1.45 |
| 1960 | 446 | 741 | 1.66 | 42 | 64 | 1.52 |
| 1970 | 470 | 754 | 1.60 | 42 | 66 | 1.57 |
| 1980 | 506 | 797 | 1.58 | 22 | 40 | 1.82 |
| 1990 | 572 | 871 | 1.52 | 22 | 26 | 1.18 |
| 2000 | 609 | 892 | 1.46 | 18 | 27 | 1.50 |

Table 27: Average number of words within the noun phrases for *stop NP from -ing* and *stop NP -ing*: frequencies in words (cf. Figure 10, Graphs 9 and 10, Section 5.3, Chapter 5)

| **Year** | ***stop NP from -ing*** | | | ***stop NP -ing*** | | |
|---|---|---|---|---|---|---|
| | **Raw frequency (instances)** | **Number of words within noun phrases** | **Average number of words (per words)** | **Raw frequency (instances)** | **Number of words within noun phrases** | **Average number of words (per words)** |
| 1810 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1820 | 2 | 4 | 2.00 | 1 | 2 | 2.00 |
| 1830 | 5 | 6 | 1.20 | 1 | 2 | 2.00 |
| 1840 | 4 | 7 | 1.75 | 2 | 3 | 1.50 |
| 1850 | 5 | 8 | 1.60 | 4 | 5 | 1.25 |
| 1860 | 4 | 6 | 1.50 | 4 | 4 | 1.00 |
| 1870 | 12 | 19 | 1.58 | 6 | 10 | 1.67 |
| 1880 | 10 | 13 | 1.30 | 12 | 21 | 1.75 |
| 1890 | 15 | 22 | 1.47 | 15 | 22 | 1.47 |
| 1900 | 25 | 32 | 1.28 | 12 | 16 | 1.60 |
| 1910 | 31 | 41 | 1.32 | 26 | 36 | 1.38 |
| 1920 | 58 | 79 | 1.36 | 24 | 33 | 1.38 |
| 1930 | 60 | 92 | 1.53 | 30 | 46 | 1.53 |
| 1940 | 69 | 100 | 1.45 | 21 | 28 | 1.33 |
| 1950 | 83 | 113 | 1.36 | 29 | 47 | 1.62 |
| 1960 | 131 | 173 | 1.32 | 37 | 55 | 1.49 |
| 1970 | 143 | 190 | 1.33 | 49 | 67 | 1.37 |
| 1980 | 171 | 263 | 1.54 | 37 | 64 | 1.73 |
| 1990 | 278 | 357 | 1.28 | 33 | 45 | 1.36 |
| 2000 | 354 | 479 | 1.35 | 35 | 43 | 1.23 |